

977476329

# CONTENTS OF VOL. XXXVII. (Nos. 264-275)

## (SEVENTH SERIES) 1946

### NUMBER 264—JANUARY

	Page
I. A Photo-Electric Fourier Transformer. By R. FÜRTH and R. W. PRINGLE, Edinburgh .....	1
II. Fatigue in Selenium Barrier Layer Photocells. By R. A. HOUSTOUN, D.Sc., University of Glasgow.....	13
III. Finite Hankel Transforms. By IAN N. SNEDDON .....	17
IV. The Measurement of Pressure on the Sea Bed. By E. G. RICHARDSON. (Plate I.) .....	25
V. Colour Vision. By JOHN H. SHAXBY, D.Sc., F.Inst.P., Physiology Institute, University College, Cardiff .....	33
VI. A Centrifugal Method of Measuring the Surface Tensions and Interfacial Tensions of Liquids. By W. MEYERSTEIN, M.D. (Department of Physiology, University of Birmingham), and J. D. MORGAN, D.Sc.....	41
VII. A Non-Quantum Indication of Planck's Radiation Formula. By HERBERT DINGLE, Professor of Natural Philosophy, Imperial College of Science and Technology .....	47
VIII. A Relativistic Determination of the Relation between Electrical Resistance and Temperature. By HERBERT DINGLE, Professor of Natural Philosophy, Imperial College of Science and Technology .....	58
IX. On the Dimensions of Physical Magnitudes (Sixth Paper). By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc. ....	64
X. Boundary Value Problems of a Heavy Circular Disk held in a Vertical Plane. By BIBHUTIBHUSAN SEN .....	66
XI. Notice respecting New Book :— 'The Advancement of Science.'—Vol. III. No. 12 (April 1946)	72

### NUMBER 265—FEBRUARY

XII. Solar Radio Noise.—I. By Sir EDWARD APPLETON, F.R.S., and J. S. HEY, M.Sc. (Plate II.) .....	73
XIII. An Expression for the Sine of a Fourier Series. By C. STRACHEY and P. J. WALLIS .....	84
XIV. Hahn's Functions $S_m(\alpha)$ and $U_m(\alpha)$ . By C. STRACHEY and P. J. WALLIS .....	87
XV. The Influence of Spray Particle Size and Distribution in the Combustion of Oil Droplets. By R. P. PROBERT .....	94
XVI. An Escalator Process for the Solution of Linear Simultaneous Equations. By J. MORRIS, B.A. ....	106
XVII. Viscosities of the Di-hydroxy Benzenes and some of their Derivatives. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES ....	120
XVIII. The Complete, Three-Dimensional Colour Domain and its Metrics. By LUDWIK SILBERSTEIN, Ph.D. Communication No. 1027 from the Kodak Research Laboratories .....	126

## NUMBER 266—MARCH

	Page
XIX. A Problem on the Summation of Simple Harmonic Functions of the same Amplitude and Frequency but of Random Phase. By F. HORNER, M.Sc. (Radio Division, National Physical Laboratory) . . . . .	145
XX. A Magnetic Lens of Special Field Form for $\beta$ - and $\gamma$ -Ray Investigations, Designs and Applications. By KAI SIEGBAHN . . . . .	162
XXI. Crystal Habit and Internal Structure.—I. By A. F. WELLS, M.A., Ph.D. . . . .	184
XXII. On a Difficulty in the Application of the Law of Mass Action. By J. A. CHALMERS, M.A., Ph.D., and C. W. GIBBY, M.Sc., Ph.D., Durham Colleges in the University of Durham . . . . .	200
XXIII. The Viscosities and Structures of Acid Chlorides and Amides. By J. NEWTON FRIEND and WILLIAM D. HARGREAVES . . . . .	201
XXIV. The Field of a Charged Semi-Infinite Rectangular Conductor Parallel to an Earthed Infinite Plane Conductor or the Flow from a Thick-Walled Jet. By N. DAVY, University College, Nottingham . . . . .	207
Errata . . . . .	216

## NUMBER 267—APRIL

XXV. Crystal Habit and Internal Structure.—II. By A. F. WELLS, M.A., Ph.D. . . . .	217
XXVI. Four-figure Tables of the Airy Function in the Complex Plane. By P. M. WOODWARD, B.A., and Mrs. A. M. WOODWARD, M.A. . . . .	236
XXVII. The Determination of the Orientation of Single Crystals on a Weissenberg Goniometer. By NORA WOOSTER, M.A., Ph.D., and W. A. WOOSTER, M.A., Ph.D., F.Inst.P., Brooklyn Crystallographic Laboratory, Cambridge . . . . .	262
XXVIII. Theory of Diffraction. By F. B. PIDDUCK . . . . .	280
XXIX. On a Relation between some Natural Constants and the Lesser Particles. By Dr. A. R. STOKES . . . . .	287

## NUMBER 268—MAY

XXX. Random Fluctuations in a Cathode Ray Oscillograph. By N. R. CAMPBELL, Sc.D., F.Inst.P., and V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P., A.M.I.E.E. . . . .	289
XXXI. The Structure of an Electromagnetic Field in the Neighbourhood of a Cusp of a Caustic. By T. PEARCEY . . . . .	311
XXXII. A Problem in Non-Steady Heat Conduction. By L. R. BROWN, Statistics and Mechanics Laboratory, the B.C.U.R.A. . . . .	318
XXXIII. The Sensitivity of a Space-Charge Detector to the Presence of Positive Ions of Helium with Small Kinetic Energies. By D. M. MILLEST, Ph.D. . . . .	323
XXXIV. An Experimental Study of the Quantum Nature of X-rays. By M. P. GIVENS, Physics Department, The Pennsylvania State College, State College, Pa. . . . .	335



	Page
XXXV. The Photoelectric Mechanism of Selenium Barrier Layer Elements. By ARNE ELD SANDSTRÖM .....	347
XXXVI. Note on the Empirical Relations between Chemical Constitution and Thermal Expansion Coefficient of Organic Liquids. By GUNNAR STEENSHOLT .....	357

## NUMBER 269—JUNE

XXXVII. Interference Fringes due to a Wedge of Air or Liquid between Two Glass Plates, in Theory and Practice. By G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge .....	361
XXXVIII. New Contributions to Interferometry.—Part VII. Monochromatic and White-light Multiple Beam Interferometric Studies on Mica. By S. TOLANSKY, Ph.D., D.I.C., Manchester University. (Plates III.–V.)	390
XXXIX. A Method for the Determination of Crystal Cuts by applying the Reflection of X-rays from a known Lattice Plane. By V. PETRŽILKA and J. BENEŠ. (Plate VI.) .....	399
XL. The Propulsive Force produced by an Intermittent Jet of Gas. By J. L. KASTNER, M.A. ....	411
XLI. Kinematic Relativity. By Professor W. WILSON, F.R.S. ....	421
XLII. Forced Vibrations of a Whirling Wire. By A. VICTOR MASKET, Ph.D., Mechanics and Electricity Division, Naval Research Laboratory, Washington, D.C. ....	426

## NUMBER 270—JULY

XLIII. Theory of the High Pressure Mercury Vapour Discharge. By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E. ....	433
XLIV. New Contributions to Interferometry.—Part VIII. Interferometric Examination of Small Local Irregularities in Mica. By S. TOLANSKY, Ph.D., D.I.C. Manchester University. (Plates VII. & VIII.)	453
XLV. A Photo-elastic Study of Stress Waves. By D. A. SENIOR, B.A., and A. A. WELLS, B.Sc., Naval Construction Research Establishment. (Plates IX.–XII.) .....	463
XLVI. On the Cause of the Red Shifts in the Spectra of the Extragalactic Nebulae. By G. J. WHITROW .....	469
XLVII. Power Laws of Latent Heat, Orthobaric Density, Surface Tension and Viscosity of Liquids. By W. J. JONES and S. T. BOWDEN ..	480
XLVIII. The Measurement of Young's Modulus and its Temperature Coefficient for Short Filaments by Flexural Vibrations. By Y. L. YOUSEF, M.Sc., Fouad I. University, Cairo .....	490
XLIX. A Bridge Method for the Investigation of Non-linear Resistors. By G. T. BAKER, M.Sc. (Circuit Engineer, Automatic Telephone and Electric Co., Ltd.) .....	498
L. Dr. Löwy's Theory of Ground-water Accumulation. By B. A. KEEN	502
LI. Notice respecting New Book :— Dr. M. H. PIRENNE's The Diffraction of X-rays and Electrons by Free Molecules .....	504

## NUMBER 271—AUGUST

	Page
LII. Further Notes and Suggestions on the Teaching of Physics. By C. J. SMITH, Ph.D., M.Sc., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey .....	505
LIII. A "Tangent-law" Photometer. By the Rev. G. D. YARNOLD, University College, Nottingham. (Plate XIII.) .....	534
LIV. Generalized Routh-Hurwitz Discriminant. (An Extension of the Theorems of Sturm, Routh and Hurwitz on the Roots of Polynomial Equations.) By SEYMOUR SHERMAN .....	537
LV. A Critical Examination of Milne's Kinematical Relativity. By WILLIAM BAND, Department of Physics, Yenching University, Peiping, China .....	551
LVI. On the Theory of Boundary Perturbations. By G. D. WASSERMANN, The Mining and Technical College, Wigan .....	563
LVII. A Kinetic Energy Correction for the Flow of Plastic Liquids through Circular Pipes. By T. R. LOMER, B.Sc. ....	571
LVIII. Partitioning of Space into Enantiomorphous Polyhedra. By SIDNEY MELMORE .....	574
LIX. Notices respecting New Books :—	
Mr. R. V. SOUTHWELL's Relaxation Methods in Theoretical Physics .....	576
Advances in Colloid Science (edited by H. MARK and G. S. WHITBY).—Vol. II. Scientific Progress in the Field of Rubber and Synthetic Elastomers .....	577
Physical Methods of Organic Chemistry (edited by A. WEISSBERGER).—Vol. II .....	577
Mr. H. HETÉNYI's Beams on Elastic Foundation .....	578

## NUMBER 272—SEPTEMBER

LX. Further Notes and Suggestions on the Teaching of Physics. By C. J. SMITH, Ph.D., M.Sc., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey .....	579
LXI. Crystal Habit and Internal Structure.—III. By A. F. WELLS. (Plate XIV.) .....	605
LXII. The Ionization of Mercury Vapour by Slow Electrons. By the late T. McFADDEN, Ph.D., Physics Department, Queen's University, Belfast .....	630
LXIII. On the Radial Error in a Gaussian Elliptical Scatter. By D. MARTIN, M.A., B.Sc. ....	636
LXIV. Scattering of X-rays of High Frequency. By N. WINOGRADOFF, Ph.D., A.Inst.P. ....	640
LXV. An Illusion of Size. By ROBERT WEIL, B.Sc., A.Inst.P., South-west Essex Technical College. ....	643
LXVI. On a Derivation of the Equations of Equilibrium of a Thin Plate. By J. G. OLDROYD, M.A., Courtaulds Ltd., Research Laboratory, Maidenhead, Berks. ....	648
LXVII. On Jeans' Dynamical Theory of Gases. By W. F. SEDGWICK ..	651



## NUMBER 273—OCTOBER

	Page
LXVIII. Theory of the High Pressure Mercury Vapour Discharge (B). By V. J. FRANCIS, B.Sc., F.Inst.P., A.M.I.E.E. ....	653
LXIX. The Limiting Resolving Power of the Electron Microscope. By G. LIEBMANN .....	677
LXX. The Effect of Fluctuation Noise Interference on Pulse Distortion. By P. J. HILTON, B.A. ....	685
LXXI. "Lagrangean" Formulae for the Direct Calculation of Harmonic Output and of Intermodulation Products. By A. BLOCH, Dr.-Ing., M.Sc. (Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England) .....	694
LXXII. Haidinger's Rings in Divergent Light. By G. N. RAMACHAN- DRAN (from the Department of Physics, Indian Institute of Science, Bangalore) .....	700
LXXIII. A Discussion of the Pressure-tendencies associated with Gradient and Horizontal Geostrophic Flow. A Formula for the Variation with Height of the Vertical Velocity. By A. G. MATTHEWMAN, M.A. ....	706
LXXIV. A Criticism of Two Well-known Theorems on Most Efficient Propellers. By V. D. NAYLOR, M.Sc. ....	717
LXXV. Determination of Refractive Index of a Liquid by Newton's Rings. By R. L. WOOLLEY, B.A., Trinity College, and G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge .....	728

## NUMBER 274—NOVEMBER

LXXVI. Plastic Stresses in a Semi-Infinite Cohesive Mass due to a Surcharge covering a Strip of the Surface of Infinite Length. By N. S. BOULTON, M.Sc., Assoc.M.Inst.C.E., University of Sheffield .....	733
LXXVII. A Method for the Measurement of the Dielectric Properties of Liquids in the Frequency Range 600–3,200 Mc./sec. (50–9.4 cm.). By R. DUNSMUIR, M.Sc. and J. G. POWLES, M.Sc., Electrotechnics Depart- ment, Victoria University, Manchester .....	747
LXXVIII. Large Deflections of Circular and Square Plates. By H. D. CONWAY, M.A., Ph.D. ....	756
LXXIX. The Large Deflections of Rectangular Membranes and Plates. By H. D. CONWAY, M.A., Ph.D. ....	767
LXXX. A Contribution to Linear Network Analysis. By D. K. C. MACDONALD, M.A., A.Inst.P., Captain, R.E.M.E., Military College of Science .....	778
LXXXI. The Luminous Efficiency of the High Pressure Mercury Vapour Discharge. By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., G. G. ISAACS, B.Sc., A.Inst.P., A.M.I.E.E., and E. H. NELSON .....	789
LXXXII. On the Outer Unipolar Rotation and the Outer Unipolar Induction. By MARIN KATALINIĆ .....	797
LXXXIII. The Thermal Ohm, Farad and Henry. By R. C. L. BOSWORTH, Ph.D., D.Sc., F.Inst.P. ....	803
LXXXIV. A Note on Fusion and the Hole-Theory of Liquids. By A. G. CHOWDRI and F. C. AULUCK (University of Delhi) .....	809
LXXXV. Notices respecting New Books :— Mr. E. A. MAXWELL's Methods of Plane Projective Geometry Based on the Use of General Homogeneous Co-ordinates. . . . . Prof. C. E. WEATHERBURN's First Course in Mathematical Statistics .....	812 812

## NUMBER 275—DECEMBER

	Page
LXXXVI. On the Bending of an Initially Straight Beam Under Arbitrary Loading. By A. W. GILLIES (Northampton Polytechnic) . . . .	813
LXXXVII. A Problem in Dynamic Balancing of Scanners for Radar. By C. Fox, M.A., D.Sc. . . . .	830
LXXXVIII. The Second Law of Thermodynamics, with Particular Reference to Carnot Theorems Relating to Regenerative, Refrigeration and Warming Cycles. By Professor W. J. WALKER, Department of Mechanical Engineering, University of the Witwatersrand, Johannesburg, South Africa . . . . .	842
LXXXIX. Mathematical Theory of Deflection of Beam. By J. G. FREEMAN, M.A., Ph.D. . . . .	855
XC. A Laboratory Investigation of Electrosmosis in Soils. By D. J. MACLEAN, B.Sc., A.Inst.P., and D. W. ROLFE, of the Road Research Laboratory, Department of Scientific and Industrial Research . . . . .	863
XCI. On the Raman Spectrum of <i>p</i> -Diphenyl Benzene. By S. K. MUKERJI, M.Sc., Ph.D. (Lond.), Professor of Physics, Agra College, Agra, and LAKSMAN SINGH, M.Sc., Agra University Research Scholar in Physics. 874 (Plate XV.)	
XCII. Notices respecting New Books :—	
Essays in Rheology (1944 Conference of the British Rheologists' Club) . . . . .	883
Dr. WRINCH's Fourier Transforms and Structure Factors . . . .	884
Index. . . . .	

## P L A T E S

- I. Illustrative of Dr. E. G. RICHARDSON's Paper on The Measurement of Pressure on the Sea Bed.
- II. Illustrative of Sir EDWARD APPLETON and J. S. HEY's Paper on Solar Radio Noise.—I.
- III.—V. Illustrative of Dr. S. TOLANSKY's Paper on New Contributions to Interferometry. Part VII.—Monochromatic and White-light Multiple Beam Interferometric Studies on Mica.
- VI. Illustrative of Messrs. V. PETRŽÍLKA and J. BENEŠ's Paper on A Method for the Determination of Crystal Cuts by applying the Reflection of X-rays from a known Lattice Plane.
- VII.—VIII. Illustrative of Dr. S. TOLANSKY's Paper on New Contributions to Interferometry. Part VIII.—Interferometric Examination of Small Local Irregularities in Mica.
- IX.—XII. Illustrative of Messrs. D. A. SENIOR and A. A. WELLS's Paper on A Photo-elastic Study of Stress Waves.
- XIII. Illustrative of the Rev. G. D. YARNOLD's Paper on A "Tangent-law" Photometer.
- XIV. Illustrative of Dr. A. F. WELLS's Paper on Crystal Habit and Internal Structure.
- XV. Illustrative of Dr. S. K. Mukerji, and Mr. Laksmee Singh on the Ramad Spectrum of *p*-Diphenyl Benzene.



THE  
PHILOSOPHICAL MAGAZINE  
A JOURNAL OF THEORETICAL EXPERIMENTAL  
AND APPLIED PHYSICS

First published in 1798

[SEVENTH SERIES—VOL. 37]

---

I. *A Photo-Electric Fourier Transformer.*

By R. FÜRTH and R. W. PRINGLE, Edinburgh\*.

[Received October 11, 1945.]

*Summary.*

A new instrument, working on a photo-electric principle, is described by which the "Fourier Transform" of a given function is automatically and instantaneously produced in the form of a graph on the screen of a cathode ray oscillograph when a mask cut out in the shape of the graph of the given function or a record of this function in density variation is inserted in the instrument.

---

1. *Purpose of the Instrument.*

THE "Fourier Transform"  $g(y)$  of a function  $f(x)$  is usually defined by the integral

$$g(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x) e^{-iyx} dx \quad (x, y \text{ real variables}). \quad \dots (1)$$

According to a well-known theorem the relation between the two functions  $f(x)$  and  $g(y)$  is reciprocal; that is, the formula

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} g(y) e^{iyx} dy \quad \dots \dots \dots (2)$$

holds for all values of  $x$  where  $f(x)$  is continuous.

---

\* Communicated by the Authors.





transparent. The above condition can always be satisfied by replacing the original function  $f(x)$  by the function

$$f'(x) = f(x) + f_0, \quad (6)$$

where  $f_0$  has a sufficiently large constant and positive value within the interval  $(a, b)$  and is zero outside that interval.

The whole area is now illuminated in such a way that the intensity of illumination varies in the  $x$ -direction according to the formula

$$J(x) = A \cos(yx + \delta) + D \quad (A, D \text{ constants}). \quad (7)$$

In other words a pattern of parallel fringes is projected on the mask in which the light intensity varies harmonically between the extreme

Fig. 1.

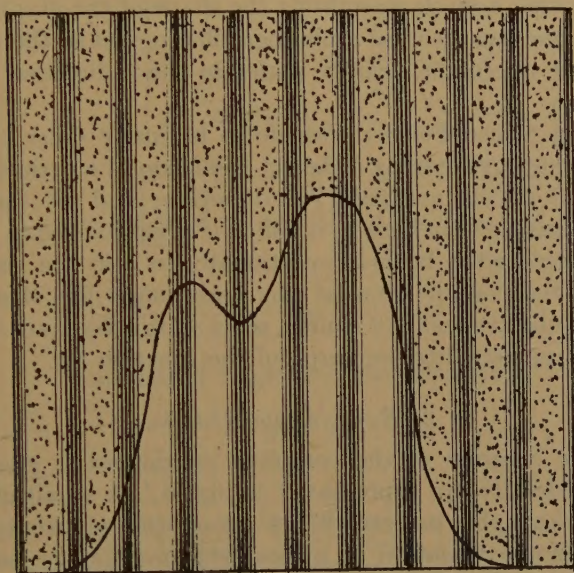


Diagram illustrating the fundamental principle of the method.

values  $(D-A)$  and  $(D+A)$ , the spacing  $\lambda$  of the pattern along the  $x$ -axis being equal to  $2\pi/y$ . We remark for later reference that the intensity at  $x=0$  is independent of  $y$  and only dependent on the value of the phase constant  $\delta$ .

If all the light which penetrates through the mask is concentrated on the cathode of a photo-electric cell the total illumination there and accordingly the photo-electric current  $i$  will, by virtue of (5), (6) and (7) (apart from a constant factor), be given by

$$\begin{aligned} i &= \int_a^b f'(x)J(x) dx = A \int_a^b f(x) \cos(yx + \delta) dx + f_0 A \int_a^b \cos(yx + \delta) dx + C \\ &= A\psi(y) + B\sigma(y) + C, \quad (8) \end{aligned}$$

where 
$$\sigma(y) = \frac{1}{y} [\sin (by + \delta) - \sin (ay + \delta)], \quad . . . . (9)$$

and where A, B, C are constants.

If  $y$  is now made to vary periodically in time between  $y=0$  and an upper limit  $y_0$ , that is, if the fringe system is made to extend and shrink periodically in time, the spacing  $\lambda$  varying between  $\lambda_0=2\pi/y_0$  and infinity, the time dependence of the photo-current  $i$  will evidently represent the function (8) within the limits  $-y_0$  and  $+y_0$ . In the particular case  $f_0=0$  (*i. e.* when the given function is everywhere positive) or for  $b=-a$  and  $\delta=\pm\pi/2$ , one has  $\sigma(y)=0$ ; hence  $i$  (apart from an additional and insignificant constant) directly represents the Fourier Transform (5) of  $f(x)$ . In all other cases the function  $\sigma(y)$  can be evaluated separately by using a mask in the shape of a rectangle with height  $f_0$  and width  $b-a$ .

Instead of using a mask resembling the shape of the graph of the given function  $f(x)$ , one can also use a photographic record of the function on a plate or a film where the density varies in the  $x$ -direction according to this function.

The plotting of the Fourier Transform is finally done by amplifying the photo-current and making it deflect the beam of a cathode ray oscillograph (C.R.O.) in the vertical (Y-) direction, and at the same time applying a harmonic time base to it which is synchronized to the variation of  $y$ , thus deflecting the beam proportionally to  $y$  in the horizontal (X-) direction. If the periodic time of this variation is less than 1/10th sec. the eye of the observer will, of course, see a stationary trace on the screen of the C.R.O. representing the graph of the function (8).

### 3. Experimental Details.

A schematic diagram of the complete instrument is given in fig. 2; some further details are represented in fig. 3. A small glass plate A, bearing a photographic pattern ( $2'' \times 2''$ ) consisting of twenty fringes and having a harmonic variation in optical transmission, is mounted inside a large journal bearing B (internal diameter  $3''$ ) so that it can be rotated with uniform angular velocity about an axis at right angles to the plate A by means of the electric motor Q (1/12 h.p., D.C.) and a belt. Fig. 4 gives a microphotogram of a section of the plate which shows that the condition (7) is reasonably well satisfied\*. A fine screw adjustment C (seen in fig. 3) allows the coincidence of any desired point of the plate with the axis of revolution.

Uniform illumination of the plate is obtained by the use of a 200-watt, 230-volt (Mazda) projector lamp D, mounted in a lamp-housing which also contains a diffuser E and a condenser lens F. Although the lamp current was taken from the rectified D.C. mains the fluctuations of light

---

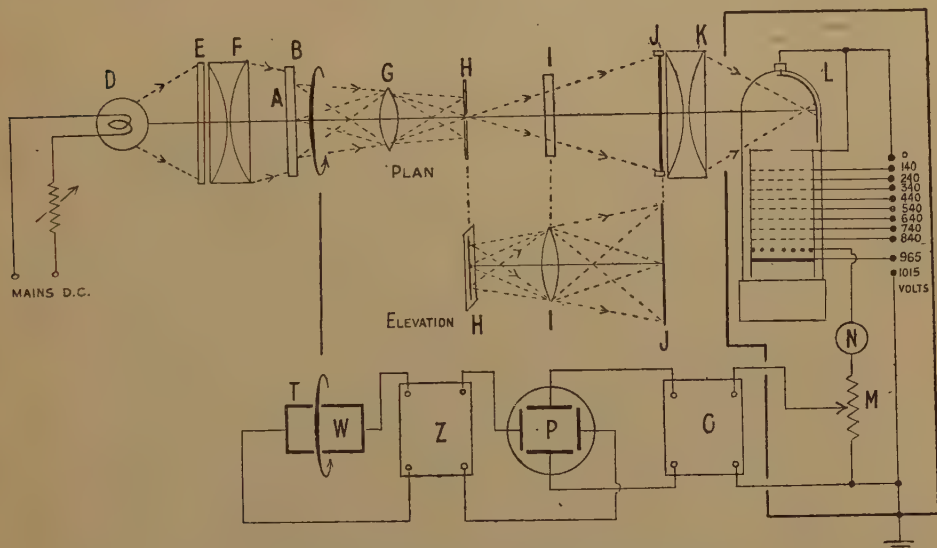
\* For the production of the pattern and the microphotogram we are indebted to Mr. W. F. Berg and Mr B. I. Davies of the Kodak Research Laboratory (Physics Department).



intensity were practically completely suppressed by the great heat capacity of the lamp filament\*.

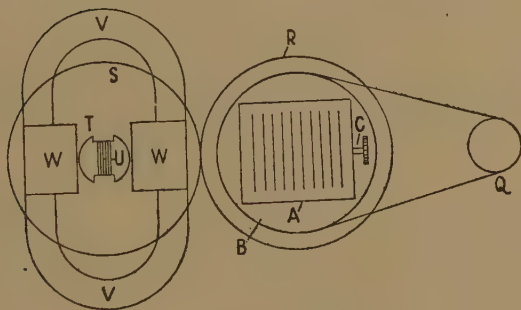
A reduced real image ( $\times 1/2$ ) of the pattern is then produced by the lens G (Zeiss Tessar photographic objective 1 : 4) on a fine adjustable vertical slit H (shown in fig. 2 in plan and elevation), which is arranged so

Fig. 2.



Schematic diagram of the instrument.

Fig. 3.



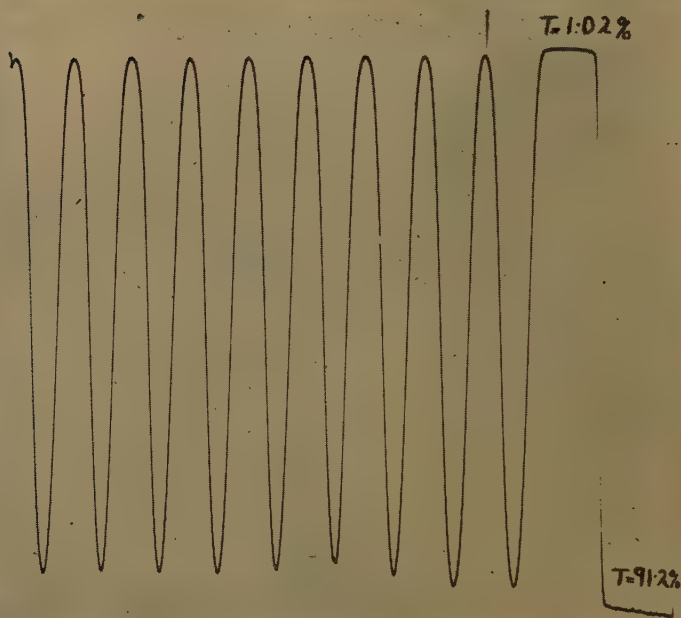
Some constructional details.

\* It was originally intended to produce the pattern by an optical interference method by placing a quartz plate between two crossed sheets of "polaroid" and using a sodium discharge lamp as a monochromatic light source. By this procedure a strictly sinoidal distribution of light intensity across the pattern is achieved, but the discharge lamp (as most probably any other sufficiently intense source of monochromatic light) gave so large fluctuations of light output that this method had to be abandoned.

as to make the axis of rotation pass exactly through the centre of the slit. In this way a distribution of light intensity along the slit is obtained which for every position of the plate A is sinusoidal provided that the slit width is small compared with  $\lambda_0$  (under the present conditions a slit width of 0.05 mm. was found suitable). Rotation of the plate A, and therefore of the pattern image, with constant angular velocity  $\omega$  causes a variation of the spacing of this light distribution according to

$$\lambda = \lambda_0 / \sin \omega t.$$

Fig. 4.



Microphotogram of the fringe pattern.

Thus the light distribution satisfies formula (7), where  $y$  varies in time according to

$$y = 2\pi/\lambda = y_0 \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

with

$$y_0 = 2\pi/\lambda_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

By means of the cylindrical lens I (with its cylinder axis at right angles to the slit) an image ( $3'' \times 3''$ ) of the slit is formed in the plane J whereby each point of the slit is broadened out into a horizontal line so that the image consists of a fringe pattern like the one shown in fig. 1, which contracts and expands in time according to the formulæ (7), (10) and



(11). By adjusting the screw C the centre of the pattern (which remains stationary) can be made bright ( $\delta=0$ ) or dark ( $\delta=\pi$ ), or to have some intermediate illumination corresponding to the particular point of the pattern disk A through which the axis of rotation passes.

A projection lantern frame with slider is fixed in the plane J, and by means of this the mask or record of the function to be analysed can be brought into position as indicated in fig. 1. The optical arrangement is completed by the condensing system K which collects the light passing through J on to the cathode of the photo-electric cell L.

As the light intensity available at the photocell is seriously reduced by the narrow slit, and as the variable part of the photo-current is only a small fraction of the total current, a photocell of very high sensitivity and very high stability of operation had to be used. Therefore, first attempts with an ordinary gas-filled photocell had to be abandoned and a Baird electron multiplier photocell type MA20 (with 9 stages of multiplication) is now used. A series of high-tension batteries insulated from one another by paraffin wax blocks and kept in a screened box supplies the potentials required for the various electrodes, as indicated in fig. 2. The photocell itself is carefully screened both electrically and optically, and is mounted on the optical bench carrying all the other components already mentioned.

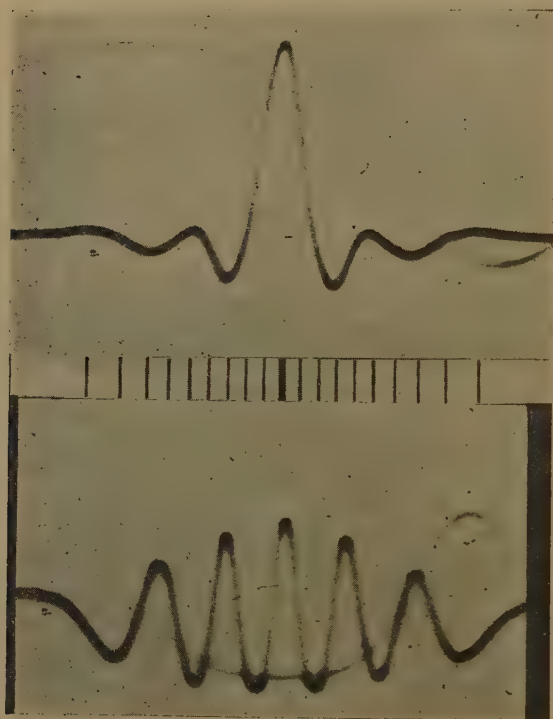
The photo-current passes through the load-resistance potentiometer M (100,000 ohms). The voltage fluctuations across M are transferred to the cathode ray oscillograph P (Mullard GM 3152) with incorporated two-stage linear amplifier O. As the current fluctuations in M are a faithful reproduction, amplified about 5,000 times, of the current output from the light sensitive cathode of L, the vertical (Y-) deflections of the oscillograph beam are precisely proportional to the variations of the light intensity and at the same time sufficiently large even when masks are used, which cover almost the whole area of J. They can be adjusted to a suitable size by the potentiometer M. The millimeter N in the photo circuit serves to ensure that the photo-current does not exceed the permitted maximum of 1 mA. In actual practice, however, the current is always much smaller.

In order to produce the graph of the Fourier Transform one has still to supply a time-base satisfying formula (10), *i.e.* to supply a voltage to the X-deflecting plates of the C.R.O., varying harmonically in time with angular frequency  $\omega$ . The following electro-magnetic method proved eventually satisfactory. A gear-wheel R (2.6" diameter, 128 teeth) was mounted on the journal bearing B, as shown in fig. 3 (the teeth are not shown in the figure), in such a way as not to interfere with the optical system. Another wheel S of equal size was mounted on a spindle consisting of two sections screwed axially from both ends into a simple H-type ebonite former T bearing a coil of 100 turns of S.W.G. 36 insulated wire. The ends of the coil are connected to the two insulated sections of the spindle, and electrical connections to them are made

through the bearing supports, aided by spring contacts pressing against the spindle.

The coil rotates in a strong and reasonably uniform magnetic field of about 1,500 oersted, produced by two Alcomax magnets V (M 3700, J. Neill & Co., Sheffield) and two iron pole pieces W (area  $1\frac{1}{2}$  square inches), which are assembled in such a way that the like poles of the magnets are facing each other. By rotating the coil with about 10 rev./sec. a sinusoidal voltage of about 1 volt R.M.S. is generated in the coil, which is first

Fig. 5.

*u**b*

Examples of records of Fourier Transforms on a linear time base.

amplified and then applied to the X-deflecting plates of the C.R.O. The amplitude and thus the scale of the graph in the X-direction can be adjusted by the gain control of the amplifier Z.

Proper synchronisation of the time base to the rotation of the disc A is achieved by the correct meshing of the gear-wheels R and S so as to make the spot on the C.R.O. screen pass through  $X=0$  for  $y=0$ , *i. e.* when the fringes of A are parallel to the slit H. Because of the slight imperfection of the pattern a slight blurring of the graph on the C.R.O.



screen is caused due to the two superposed traces which are produced during each revolution of the disk. In order to remove this imperfection a paxoline disk (not shown in fig. 3) is mounted on the spindle of the generator, cut in such a way as to interrupt the light beam during one half of the cycle.

Instead of using the harmonic time base generator one can, of course, also use a linear time base such as provided by the incorporated time base of the C.R.O., which can be synchronised to the rotation of the disk. In this case, however, the scale of the abscissæ in the graph is not linear, the relation between  $X$  and  $y$  being

$$X = X_0 \arcsin y/y_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

which makes the use of the graphs a little inconvenient. No selector is necessary in this case as no superposition of traces is taking place, but on the other hand the "fly back" of the linear time base cuts through the graph and, unless a special synchronising device is used, the trace cannot be held stationary on the screen over a long period.

#### 4. Application of the Instrument.

The simplest procedure to get permanent records of the Fourier Transforms is to take photographs of the traces on the C.R.O. screen. The photographs, which are reproduced as examples of the performance of the Fourier Transformer in figs. 5 and 6, were taken with a Zeiss Contax camera with 1:2 Sonnar lens (and an additional extension piece) on 35 mm. Kodak Panatomic-X-Film, with 1/5th sec. exposure time. Enlargements of the negatives were made on Ilford Ultra Contrasty Bromide paper to a convenient size, namely so as to make the total horizontal range of the graphs  $2y_0 = 10$  cm. The two records of fig. 5 were made using the linear time base of the C.R.O. The correct  $y$ -scale is printed between the graphs. The records of fig. 6 were taken with the harmonic time base in operation and thus have a uniform  $y$ -scale.

The replacing of one function mask (or record) by another, the setting of the phase by adjusting the screw C, the adjustment of the X- and Y-scale, and the photographing of the trace takes only a couple of minutes, so that a large number of Fourier Transform graphs can be obtained in a short time.

The Y-scale unit of the graphs is, of course, arbitrary and is of no significance, as it is only the relative ordinates of the Fourier Transforms that matter. The X-scale unit, however, must be exactly known as it determines the "frequency" scale of the continuous Fourier spectrum. It is convenient to introduce instead of  $x$  and  $y$  the dimensionless variables

$$\xi = x/\lambda_0, \quad \eta = y/y_0 = y \cdot \lambda_0/2\pi. \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Substituting these variables into the expression (5) and putting

$$f(x) = \phi(\xi), \quad \psi(y) = \frac{1}{\lambda_0} \chi(\eta) \quad . \quad . \quad . \quad . \quad . \quad (14)$$

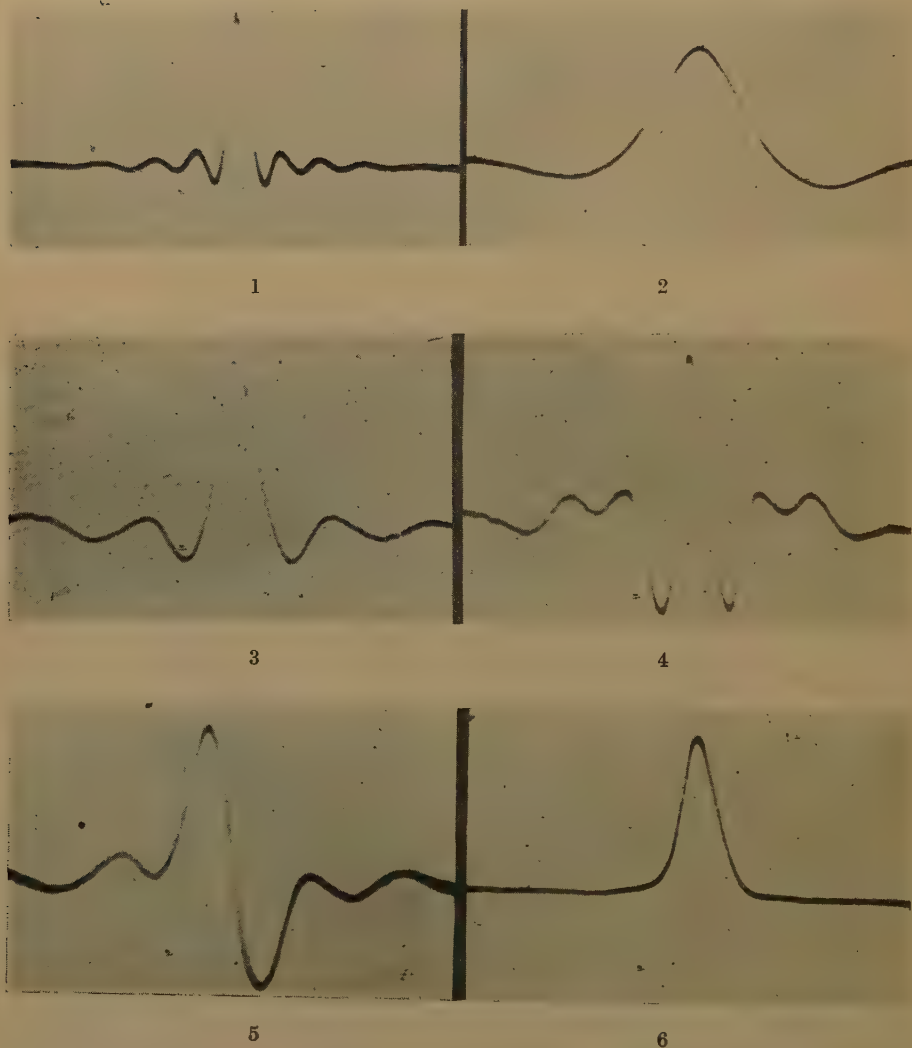
we obtain

$$\chi(\eta) = \int_{\alpha}^{\beta} \phi(\xi) \cos(2\pi\xi\eta + \delta) d\xi, \quad . . . . . (15)$$

where

$$\alpha = a/\lambda_0, \quad \beta = b/\lambda_0. \quad . . . . . (16)$$

Fig. 6.



Examples of records of Fourier Transforms on a harmonic time base.

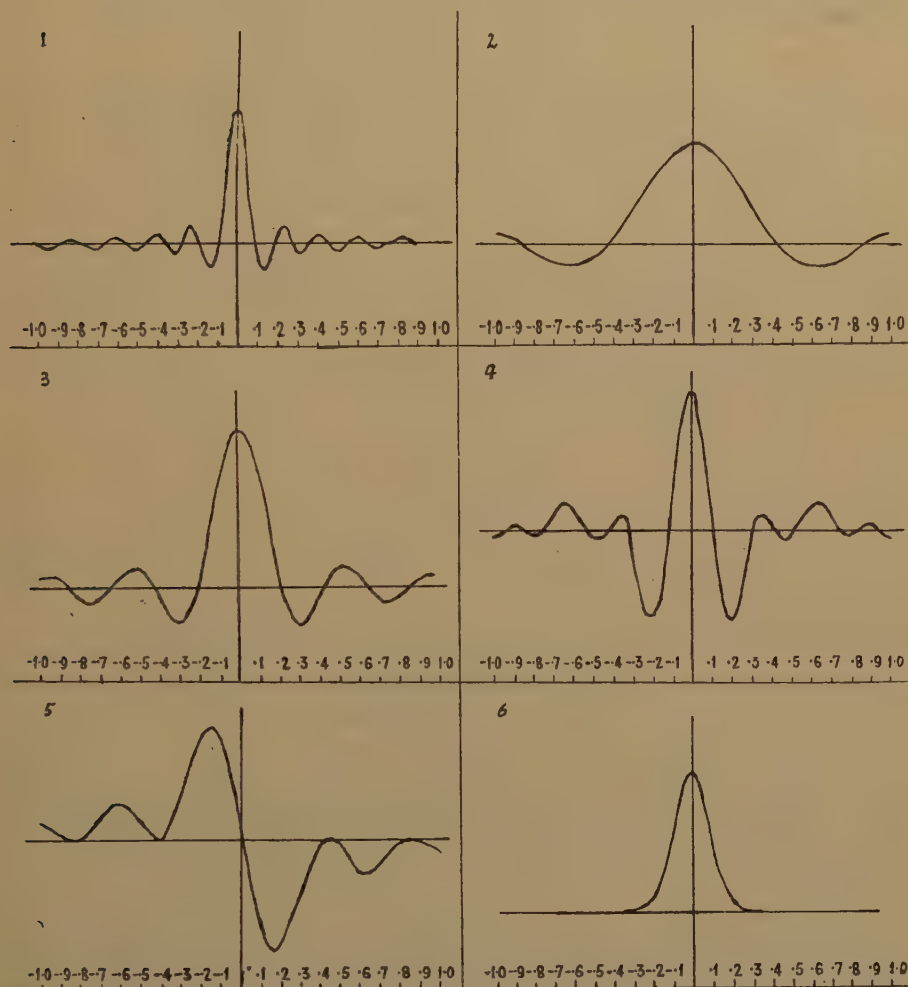
According to (10) and (11)  $\eta$  ranges from  $-1$  to  $+1$ , and this scale has been used throughout for the abscissæ of figs. 5, 6, and 7. The only instrument constant required for transforming the given function  $f(x)$  into the function  $\phi(\xi)$  appearing in the integral (15) is the minimum



spacing of the fringe pattern at the plane J, which in the present arrangement is  $\lambda_0 = 6.4$  mm. The limits of  $\xi$  imposed by the finite width of the frame for the masks are  $\pm 5.5$ .

In order to test the degree of accuracy of the present instrument the six Fourier Transforms shown in fig. 6 were calculated numerically from

Fig. 7.



Calculated graphs of the functions shown in fig. 6.

formula (15), and the curves thus obtained are reproduced in fig. 7, in equal size as those of fig. 6.

The graphs 1-5 are the Fourier Transforms of functions which are constant and positive within an interval  $(\alpha, \beta)$  and zero outside this interval. Graphs 1-4 in particular refer to  $\delta=0$  and thus represent Fourier Transforms in terms of cosine waves, and are therefore symmetric

in character. From formula (15) we obtain in these cases

$$\chi(\eta) = \text{const.} [\beta \cdot S(2\pi\beta\eta) - \alpha \cdot S(2\pi\alpha\eta)], \quad . \quad . \quad . \quad (17)$$

where

$$S(z) = \frac{\sin z}{z}. \quad . \quad . \quad . \quad (18)$$

Graph 5 refers to  $\delta = \pi/2$ , representing a Fourier Transform in terms of sine waves, and is therefore antisymmetric in character. It is easy to see that formula (17) also holds in this case when  $S(z)$  is replaced by

$$S'(z) = \frac{1 - \cos z}{z}. \quad . \quad . \quad . \quad (19)$$

Graph 1 was obtained with the full area of the frame uncovered, *i. e.*  $\alpha = -\beta = 5.5$  or  $\chi(\eta) \approx S(11\pi\eta)$ . Graph 2 was obtained by using a mask in the shape of a rectangular slit of 15 mm. width in a central position, *i. e.* for  $\alpha = -\beta = 7.5/6.4 = 1.28$  or  $\chi(\eta) \approx S(2.35\pi\eta)$ . Graphs 3 and 4 were taken using the same slit but in asymmetrical positions, namely  $\beta = 2.35$ ,  $\alpha = 0$ , and  $\beta = 3.63$ ,  $\alpha = 1.28$  resp., corresponding to the functions  $\chi(\eta) \approx S(4.7\pi\eta)$  and  $\chi(\eta) \approx \frac{1}{2}\{3S(7.05\pi\eta) + S(2.35\pi\eta)\}$ . Graph 5 belongs to the same function  $\phi(\xi)$  as Graph 3, but as already mentioned is a sine transform satisfying the formula  $\chi(\eta) \approx S'(4.7\pi\eta)$ .

Graph 6 is the cosine Fourier Transform of the "Gauss function"  $f(x) = e^{-\frac{c^2}{2}x^2}$ , which by means of (13) and (14) is transformed into  $\phi(\xi) = e^{-\frac{\gamma^2}{2}\xi^2}$  with  $\gamma = c\lambda_0$ . From (15) one obtains easily

$$\chi(\eta) = \text{const.} e^{-\frac{2\pi^2}{\gamma^2}\eta^2}, \quad . \quad . \quad . \quad (20)$$

*i. e.* as is well known, again a Gauss function. A mask in the shape of  $f(x)$  with  $c^2 = 0.007 \text{ mm.}^{-2}$  was cut out for producing the Transform (20), with  $\gamma^2 = 0.007 \cdot 6.4^2 = 0.29$ , *i. e.*  $\chi(\eta) \approx e^{-68.4\eta^2}$ .

The upper graph of fig. 5 is the same function as graph 3 of fig. 6 but taken with the linear time base, as mentioned before, in order to show the distortion produced by the non-linearity of the  $\eta$ -scale. The lower graph of fig. 5 was taken with a mask in the shape of a narrow slit, about 1 mm. wide, in an unsymmetric position and with  $\delta = 0$  which, as can be easily derived from (17), produces a modulated cosine wave resembling a "beat."

The small deviations between the curves produced by the instrument and the calculated ones are due to the imperfections of the fringe pattern plate, the slit and the optical system, and also to the time base not being perfectly sinusoidal. However, there is little doubt that further technical development of the instrument will make possible the automatic production of Fourier Transforms with any reasonably desired accuracy.

It may be finally remarked that the device could be also used for producing "Transforms" with respect to other suitable generating functions of the form  $K(xy)$  according to

$$g(y) = \int_{-\infty}^{+\infty} f(x)K(xy) dx \quad . \quad . \quad . \quad (21)$$



by replacing the pattern disk A by another one carrying a pattern of the function  $K(x)$  in density variation.

The authors wish to thank in the first place Professor M. Born for his contributions to the development of the principles of the new instrument and for his interest in the experimental work. They also thank the members of the staff of the Physics Department for apparatus lent and Mr. Mollison of the Physics Workshop for help rendered in the construction of the instrument, and finally the University of Edinburgh for a grant from the Moray Fund for the buying of apparatus.

Department of Mathematical Physics,  
University of Edinburgh.

## II. *Fatigue in Selenium Barrier Layer Photocells.*

By R. A. HOUSTOUN, D.Sc., University of Glasgow\*.

[Received October 11, 1945.]

IN a former paper † I gave some observations on the fatigue or drift of selenium barrier layer photocells. Since then I have found a better way of treating the theory of these cells ‡ and an explanation of the fatigue has been put forward by J. S. Preston §, as a consequence of which I return here to the subject.

I have found that in the selenium barrier layer cell the current is connected with the incident light by the formula

$$\frac{1}{I} = \frac{k}{E} \left( 1 + \frac{R}{r} \right) \frac{1}{L} + \frac{R}{E},$$

where  $I$  is the current,  $L$  is the intensity of the light, and the other symbols represent quantities characteristic of the cell. The photronic cell investigated at length in the paper cited above shows fatigue in a very high degree, and it seemed desirable to find how the constants of the formula were affected by the fatigue.

For this purpose, a long photometer bench was used. A 100-watt lamp was set up at one end, the photocell moved along it, and the current read at various positions. The voltage on the lamp was read by an accurate voltmeter and the readings corrected for variation of the voltage. The corrections were of the order of one per cent. The current was read

\* Communicated by the Author.

† Phil. Mag. xxxi. p. 498 (1941).

‡ Phil. Mag. xxxiii. p. 226 (1942).

§ 'Nature,' cliii. p. 681 (1944).

by a microammeter of resistance 39.4 ohms and at first this was the only external resistance in the circuit. The results are shown by the lower straight line in fig. 1, the abscissæ being measured in (foot candles) $^{-1}$  and the ordinates in  $(\mu A)^{-1}$ . The straight line, as was to be expected,

Fig. 1.

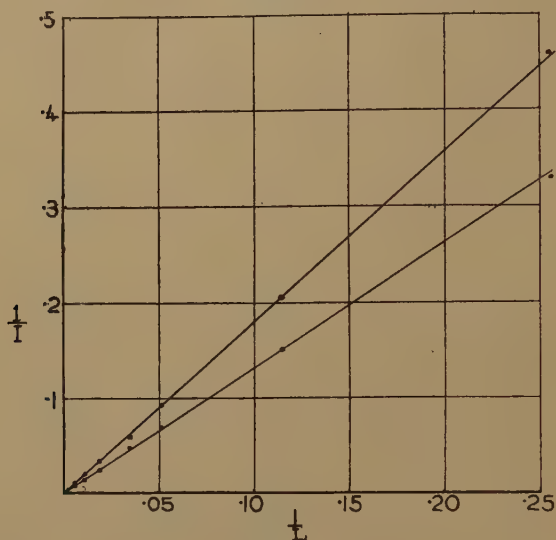
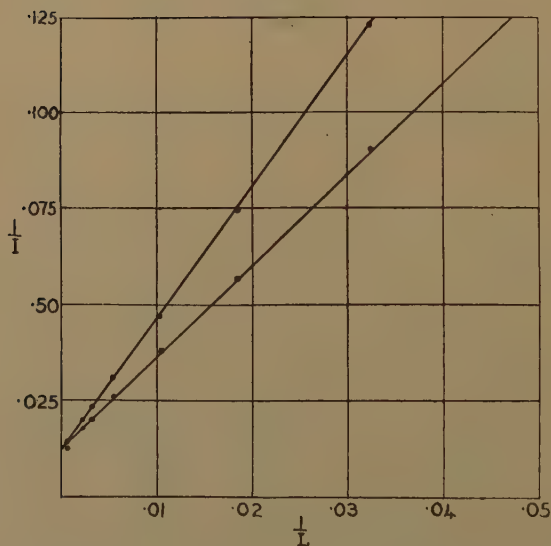


Fig. 2.



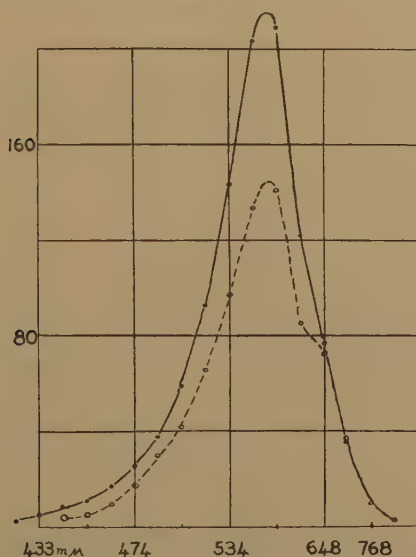
goes through the origin, the resistance  $R$  being negligible. The photocell was then fatigued by placing it close up to the 100-watt lamp for fifteen minutes and the readings repeated. The results are shown by the upper line in fig. 1,



A resistance of 2,000 ohms was then put in the circuit and, when the cell had recovered, a set of readings was taken. The results are shown by the lower line in fig. 2. The cell was then fatigued in the same way as before and a set of readings taken. These are shown by the upper line in fig. 2.

The change manifests itself in both cases as a change of gradient. There is no appreciable change in the intercept  $R/E$ . The ratio of the gradients in fig. 2 to the gradients in fig. 1 is  $1 + R/r$ , where  $r$  is the internal resistance of the photocell. If, what is sufficiently accurate, we take  $R$  as 0 in fig. 1 and as 2,000 ohms in fig. 2, we find that, when the photocell is fatigued,  $r$  goes down from 2,400 to 2,180 ohms and  $k$  goes up in the ratio 1.32 to 1.79. Both these changes act in the same direction. So the

Fig. 3



fatigue is to be regarded as a diminution of the internal resistance of the photocell together with a diminution in its sensitivity to light.

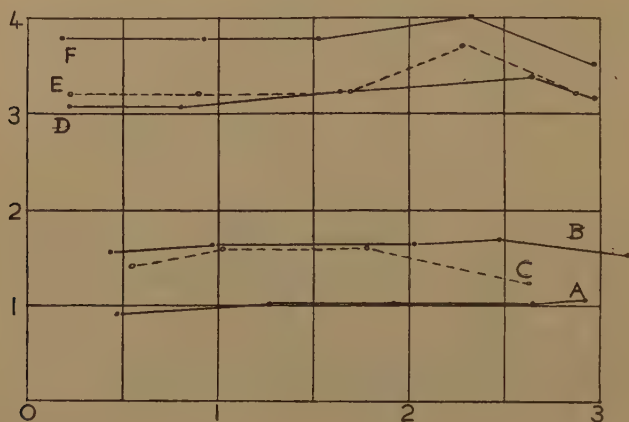
The variation of the fatigue throughout the spectrum was next investigated. The source of light was a projector lamp with its filament turned edgeways and the spectrum was produced by two flint glass prisms. The current was read by a moving coil galvanometer. The results are given by the full curve in fig. 3, where the abscissæ are deviations with some wave-lengths marked on, and the ordinates are galvanometer deflections.

The photocell was then fatigued by exposing it close up to a 100-watt lamp for 15 minutes and the observations repeated. They are represented by the broken curve in fig. 3, and it is obvious that there is no fatigue in the red, but that elsewhere it is roughly uniform all over. This is in agreement with the results of the former paper. We might explain the

result by assuming that the photoelectrons have different periods and that some are liable to fatigue and others are not.

J. S. Preston has suggested that the fatigue is due to something akin to space charge, though he does not use this expression. In an E.E.L. cell he investigated there were changes about  $\lambda=620\text{ m}\mu$ , which might be explained by the greater penetration of the light of wave-length greater than this and the consequent formation of a space charge at a greater depth in the selenium layer. A space charge should be expelled by an e.m.f. I fatigued the photocell in the usual way by applying it to the 100-watt lamp for 15 minutes; only that part of the surface on which the monochromatic radiation fell, *i. e.*, about one-quarter of its area, was exposed and the current produced was about  $\frac{1}{2}\text{ mA}$ . The variation of the current through the spectrum was then recorded. A current of about  $2\text{ mA}$  was then passed through the cell from a battery for about seven

Fig. 4.



minutes in order to drive the hypothetical space charge out and the variation of the current through the spectrum again recorded. There was no change whatever in the curve; it had exactly the same shape as the lower curve in fig. 3. Thus, so far as this cell is concerned, space charge seems to have nothing to do with the fatigue. But it may be the cause of the effect in other cells and it is interesting to note that the break in the curve in fig. 3 takes place between points at  $625$  and  $648\text{ m}\mu$ , which are just outside  $620\text{ m}\mu$ , the limit mentioned by Preston.

In the course of this investigation, readings were taken through a red glass and through a green glass for six photocells over a range of intensity varying 1,000 times. The cells were placed at various positions on a long photometer bench, and the current read, first with the red and then with the green glass in front of them, the source of light being a 100-watt lamp. A moving-coil galvanometer of resistance  $159.4$  ohms, provided with shunts, was used, with a series resistance to keep the external resistance in the circuit constant at this value. Variation of the voltage on the



lamp was corrected for. The results are shown in fig. 4 as ratio of the deflection for green to deflection for red, and the abscissæ are logarithms of the current produced by red light expressed in  $10^{-7}$  amperes. The maximum current is thus about  $100 \mu\text{A}$ . The results were consistent to about one per cent.

The cells were as follows :—A and B, Weston Photronics ; C, a cell supplied by Renny and Wender, Ltd. ; E, one of German origin ; D and F, cells supplied by Electroselenium, Ltd.

The interesting point brought out by fig. 4 is that the cells fall into two classes. The three high resistance ones, D, E, F, are somewhat red blind and A, B and C are somewhat green blind. The ratio is fairly constant over the wide range of intensity and the cells show no Purkinje effect.

### III. Finite Hankel Transforms.

By IAN N. SNEDDON\*.

[Received April 24, 1945.]

#### § 1. Introduction.

THE method of integral transforms has been successfully applied in recent years to the solution of boundary value and initial value problems in applied mathematics and is well known. In problems in which the range of values which can be assumed by one of the variables,  $x$  say, is  $(0, \infty)$  the use of an integral transform of the type

$$\int_0^\infty f(x)K(\xi_i; x) dx \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

will often reduce a partial differential equation in  $n$  independent variables to one in  $n-1$  independent variables, thus reducing the difficulty of the boundary value problem under discussion. In some instances successive operations of this type ultimately reduce the problem to a boundary value problem for an ordinary differential equation, the theory of which is far more extensively developed.

There are two obvious ways in which this method can be extended. In the first place, by extending the types of kernel  $K(\xi_i; x)$  beyond those at present in common use it may be possible to deal with boundary value problems not hitherto tractable to this method ; however, the choice of the kernel is largely determined by the form of the differential equation to be solved. The other alternative is to attempt to extend

\* Communicated by the Author.

the method to *finite* intervals, that is, to employ transforms such that the field of variation of the variable  $x$  is now  $(a, b)$ , where both  $a$  and  $b$  are real and finite. This was first suggested by Doetsch<sup>(1)</sup> in the case of Fourier kernels; defining finite Fourier transforms by the equations

$$\mathfrak{F}_n\{f(x)\} = \int_0^\pi f(x) \cos nx \, dx = \bar{f}_c(n),$$

$$\mathfrak{G}_n\{g(x)\} = \int_0^\pi g(x) \sin nx \, dx = \bar{g}_s(n).$$

Doetsch pointed out that the inversion formulæ

$$\mathfrak{F}^{-1}\{f_c\} = \frac{f_c(0)}{\pi} + \frac{2}{\pi} \sum_{n=1}^{\infty} \bar{f}_c(n) \cos nx = f(x),$$

$$\mathfrak{G}^{-1}\{g_s\} = \frac{2}{\pi} \sum_{n=1}^{\infty} \bar{g}_s(n) \sin nx = g(x),$$

follow immediately from the well-known theorems concerning Fourier series. This method has been developed by Kneitz<sup>(2)</sup> and applied recently, by Brown<sup>(3)</sup>, to some famous problems of mathematical physics. The use of transforms of this type does not solve problems which are incapable of solution by the direct method of Fourier series, but it does facilitate the resolution of boundary value problems.

The method can similarly be extended in the case of Hankel transforms—in which the kernel  $K(\xi_i; x)$  of equation (1) is a solution of Bessel's equation. A finite Hankel transform of simple type (for there are several, depending on the choice of the parameter  $\xi_i$ ) was introduced in a recent paper<sup>(4)</sup> and used there to facilitate the solution of certain initial value and boundary value problems in the theory of elastic vibrations. In the present paper an account is given of some of the simpler properties of finite Hankel transforms and of their use in solving boundary value problems for systems possessing axial symmetry. The usual method<sup>(5)</sup> of attacking such problems is to eliminate the time variable by means of a Laplace transform; this method is long and often involves the evaluation of intricate contour integrals, whereas the method of finite Hankel transforms outlined here is quick and simple to use and it avoids the use of contour integration.

## § 2. The Finite Hankel Transform.

The finite Hankel transformation is defined by the linear functional operator

$$\mathfrak{H}_\mu\{f(x)\} = \int_0^a x f(x) J_\mu(\xi_i x) \, dx = \bar{f}_J(\xi_i), \quad . \quad . \quad . \quad . \quad (2)$$

valid for all  $f(x)$  integrable in the interval  $0 \leq x \leq a$ . The set of parameters  $\xi_i$  can be chosen in more than one way, the choice being determined by the form of the inversion theorem used. If the quantities



$\xi_i$  ( $i=1, 2, 3, \dots$ ) are chosen to be the positive roots of the transcendental equation

$$J_\mu(a\xi_i)=0, \quad \dots \quad (3)$$

regarded as an equation in  $\xi_i$ , then the following inversion formula, justified by the classical theorems of Fourier-Bessel series<sup>(6)</sup>, may be used to express  $f(x)$  in terms of the set of functions  $\bar{f}_J$ :

$$f(x)=3^{-1}\{\bar{f}_J\}=\frac{2}{a^2}\sum_i \bar{f}_J(\xi_i) \frac{J_\mu(x\xi_i)}{\{J'_\mu(a\xi_i)\}^2}, \quad \dots \quad (4)$$

the summation extending over all the positive roots of equation (3).

On the other hand, if the parameters  $\xi_i$  are roots of the equation

$$\xi_i J'_\mu(\xi_i a) + h J_\mu(\xi_i a) = 0, \quad \dots \quad (5)$$

then in place of equation (4) we have the inversion formula

$$f(x)=3^{-1}\{\bar{f}_J\}=\frac{2}{a^3}\sum_i \frac{\xi_i^2 \bar{f}_J(\xi_i)}{h^2 + (\xi_i^2 - \mu^2/a^2)} \cdot \frac{J_\mu(\xi_i x)}{\{J_\mu(\xi_i a)\}^2} \quad \dots \quad (6)$$

If the field of variation of the variable  $x$  is an interval  $(a, b)$  which does *not* include the origin, then we make use of a second type of transform defined by the relation

$$\begin{aligned} \bar{\mathfrak{F}}_\mu\{f(x)\} &= \int_b^a x f(x) \{J_\mu(\xi_i x) G_\mu(\xi_i a) - G_\mu(\xi_i x) J_\mu(\xi_i a)\} dx \\ &= \bar{f}_H(\xi_i), \quad (b > a). \quad \dots \quad (7) \end{aligned}$$

In this equation  $G_\mu$  is the Bessel function of the second kind defined by the equation

$$G_\mu(z) = \frac{1}{2}\pi \operatorname{cosec} \mu z \{J_{-\mu}(z) - e^{-i\mu\pi} J_\mu(z)\}.$$

For this type of transform the appropriate inversion formula is

$$f(x) = \bar{\mathfrak{F}}_\mu^{-1}\{\bar{f}_H\} = \sum_i 2\xi_i^2 \frac{\bar{f}_H(\xi_i) J_\mu^2(\xi_i b)}{J_\mu^2(\xi_i a) - J_\mu^2(\xi_i b)} \{J_\mu(\xi_i x) G_\mu(\xi_i a) - G_\mu(\xi_i x) J_\mu(\xi_i a)\}, \quad \dots \quad (8)$$

but now the sum is taken over all the positive roots of the equation

$$J_\mu(\xi_i b) G_\mu(\xi_i a) - G_\mu(\xi_i b) J_\mu(\xi_i a) = 0. \quad \dots \quad (9)$$

It is clear that no result can be obtained by the use of finite Hankel transforms of the types defined above, which could not have been established by employing directly the well-known properties of Fourier-Bessel expansions; the sole advantage to be derived from the use of these forms of the Fourier-Bessel expansion theorems is, as will be seen from the examples below, that it greatly reduces the amount of *formal* calculation involved in the solution of certain types of initial value and boundary value problems.

### § 3. Properties of Finite Hankel Transforms.

(i) In the problems which follow we shall require some simple properties of the finite Hankel transform defined by equation (2); these are given

below. The proofs follow directly from the definition (2) and from familiar properties of the Bessel function  $J_\mu(\xi_i x)$ .

(i a) In the first instance we suppose that  $\xi_i$  is a positive root of equation (2); then

$$\mathfrak{I}_\mu(x^{\mu+1}) = a^{\mu+1} J_{\mu+1}(a\xi_i)/\xi_i, \quad \dots \quad (10)$$

$$\mathfrak{I}_0(a^2 - x^2) = 4a J_1(a\xi_i)/\xi_i^3, \quad \dots \quad (11)$$

$$\mathfrak{I}_0\left\{\frac{J_0(ax)}{J_0(a\xi_i)} - 1\right\} = -\frac{a J_1(a\xi_i)}{\xi_i(1 - \xi_i^2/a^2)}, \quad \dots \quad (12)$$

$$\mathfrak{I}_\mu\left\{x^{-1} \frac{\partial f}{\partial x}\right\} = \frac{1}{2} \xi_i [\mathfrak{I}_{\mu+1}\{f(x)\} - \mathfrak{I}_{\mu-1}\{f(x)\}] \quad \mu \neq 0, \quad \dots \quad (13)$$

$$\mathfrak{I}_0\left\{x^{-1} \frac{\partial f}{\partial x}\right\} = f(0) + \xi_i \mathfrak{I}_1\{f(x)\}, \quad \dots \quad (14)$$

$$\mathfrak{I}_\mu\left\{\frac{\partial^2 f}{\partial x^2} + \frac{1}{x} \frac{\partial f}{\partial x}\right\} = -\xi_i^2 \mathfrak{I}_\mu\{f(x)\} + \mu^2 \mathfrak{I}_\mu\left\{\frac{1}{x^2} f(x)\right\} - a \xi_i f(a) J'_\mu(\xi_i a). \quad \dots \quad (15)$$

If  $\mu=0$  and  $f(a)=0$  then (15) reduces to

$$\mathfrak{I}_0\left\{\frac{\partial^2 f}{\partial x^2} + \frac{1}{x} \frac{\partial f}{\partial x}\right\} = -\xi_i^2 \mathfrak{I}_0\{f(x)\} = -\xi_i^2 \bar{f}_J. \quad \dots \quad (16)$$

(i b) Similar results hold for the transform (2) when the set of parameters  $\xi_i$  are the positive roots of equation (5); the only property we shall make use of is:—

$$\mathfrak{I}_0\left\{\frac{\partial^2 f}{\partial x^2} + \frac{1}{x} \frac{\partial f}{\partial x}\right\} = -\xi_i^2 \bar{f}_J + a J_0(\xi_i a) \left[\frac{\partial f}{\partial x} + hf\right]_{x=a}. \quad \dots \quad (17)$$

(ii) Similarly from the definition (7) and the well-known result

$$J'_\mu(\xi_i b) G_\mu(\xi_i a) - G'_\mu(\xi_i b) J_\mu(\xi_i a) = \frac{1}{\xi_i b} \cdot \frac{J_\mu(\xi_i a)}{J_\mu(\xi_i b)}$$

(where the  $\xi_i$  are roots of equation (9)), we have

$$\mathfrak{I}_\mu\left\{\frac{\partial^2 f}{\partial x^2} + \frac{1}{x} \frac{\partial f}{\partial x} - \frac{\mu^2}{x^2}\right\} = -\xi_i^2 \mathfrak{I}_\mu(f) + \left[\frac{a}{b} f(a) - f(b)\right] \frac{J_\mu(\xi_i a)}{J_\mu(\xi_i b)}, \quad \dots \quad (18)$$

and

$$\mathfrak{I}_1\left\{\frac{x^2 - a^2}{x}\right\} = -\frac{b^2 - a^2}{b \xi_i^2} \cdot \frac{J_1(\xi_i a)}{J_1(\xi_i b)}. \quad \dots \quad (19)$$

#### §4. The Vibrations of a Circular Membrane.

As an example of the use of the finite Hankel transform (2) we shall consider the vibrations of a thin, perfectly flexible, material lamina in the form of a circle of radius  $a$  and of uniform surface density  $\sigma$  throughout. If the membrane is stretched by a tension  $T$  and if a uniform symmetrical pressure  $p(r, t)$  is applied to its surface, then the equation governing the motion of a point  $(x, y, z)$  of the membrane is

$$\frac{\partial^2 z}{\partial r^2} + \frac{1}{r} \cdot \frac{\partial z}{\partial r} = \frac{1}{c^2} \frac{\partial^2 z}{\partial t^2} - \frac{p(r, t)}{T}, \quad \dots \quad (20)$$



where  $(x, y, 0)$  are the co-ordinates of the point in the equilibrium position,  $r^2 = x^2 + y^2$  and  $c^2 = T/\sigma$ .

Multiplying equation (20) throughout by  $rJ_0(\xi_i r)$  and integrating with respect to  $r$  from 0 to  $a$  we have, by equation (16) and the boundary condition  $z=0$  when  $r=a$ , that the finite Hankel transform of order  $\mu=0$ ,

$$\bar{z}_J = \int_0^a r J_0(\xi_i r) z(r, t) dr, \quad \dots \quad (21)$$

satisfies the ordinary linear differential equation

$$\left( \frac{d^2}{dt^2} + c^2 \xi_i^2 \right) \bar{z}_J = \frac{1}{\sigma} \bar{p}_J(\xi_i, t), \quad \dots \quad (22)$$

where  $\xi_i$  is a root of equation (3) and  $\bar{p}_J$  denotes the finite Hankel transform of  $p(r, t)$  defined by an equation similar to (21).

When the function  $p(r, t)$  is known we can find the solution of equation (22) and obtain  $z$  by means of the inversion formula (4).

(a) *Free Vibrations*.—If the vibrations of the membrane are free we may take  $\bar{p}_J$  in equation (22) to be zero. Consider the membrane to be set in motion from the position  $z=f(r)$  with a velocity  $\dot{z}=g(r)$ . Then at  $t=0$  we have the initial values

$$\bar{z}_J = \int_0^a \eta f(\eta) J_0(\xi_i \eta) d\eta, \quad \frac{d\bar{z}_J}{dt} = \int_0^a \eta g(\eta) J_0(\xi_i \eta) d\eta,$$

so that the appropriate solution of equation (22) is

$$\bar{z} = \cos c\xi_i t \int_0^a \eta f(\eta) J_0(\xi_i \eta) d\eta + \frac{\sin c\xi_i t}{c\xi_i} \int_0^a \eta g(\eta) J_0(\xi_i \eta) d\eta.$$

Inverting by equation (4) we obtain finally for the displacement of the membrane

$$z = \frac{2}{a^2} \sum_i \frac{\cos c\xi_i t J_0(\xi_i r)}{[J_1(\xi_i a)]^2} \int_0^a \eta f(\eta) J_0(\xi_i \eta) d\eta \\ + \frac{2}{ca^2} \sum_i \frac{\sin c\xi_i t}{\xi_i} \cdot \frac{J_0(\xi_i r)}{[J_1(\xi_i a)]^2} \int_0^a \eta g(\eta) J_0(\xi_i \eta) d\eta, \quad \dots \quad (23)$$

where the summation with respect to  $i$  extends over all the positive roots of equation (3).

If the motion is started from the equilibrium position we may put  $f(r) \equiv 0$  in equation (23) to obtain a result which has been established by the use of the Laplace transform<sup>(7)</sup>.

(b) *Forced Vibrations*.—To illustrate the use of this method in the discussion of the forced vibrations of a circular membrane we consider forced vibrations of a simple type: suppose a membrane of radius  $a$ , stretched by a tension  $T$ , is at rest in its equilibrium position; at time  $t=0$  a pressure  $P_0\psi(t)$  is applied uniformly to the part of the surface of the membrane bounded by the circle  $r \leq b \leq a$ . Then, by equation (10),

$$\bar{p}_J = b J_1(b\xi_i) P_0\psi(t) \xi_i^{-1}$$

and equation (22) becomes

$$\left(\frac{d^2}{dt^2} + c^2 \xi_i^2\right) \bar{z}_J = \frac{P_0}{\sigma} \cdot \frac{b J_1(b \xi_i)}{\xi_i} \psi(t). \quad (24)$$

If the membrane is set in motion from rest in its equilibrium position  $\bar{z}_J = d\bar{z}_J/dt = 0$ , when  $t=0$ , so that the solution of equation (22) is

$$\bar{z}_J = \frac{P_0}{\sigma} \cdot \frac{b J_1(b \xi_i)}{c \xi_i^2} \int_0^t \psi(\eta) \sin c \xi_i(t-\eta) d\eta,$$

whence, by equation (4),

$$z = \frac{2P_0 b}{c \sigma a^2} \sum_i \frac{J_1(b \xi_i) J_0(\xi_i)}{\xi_i^2 [J_1(\xi_i a)]^2} \int_0^t \psi(\eta) \sin c \xi_i(t-\eta) d\eta, \quad (25)$$

where, as before, the sum is taken over all the positive roots of the equation (3).

There are two cases of special interest: when  $a=b$  and when  $b=0$ . If  $a=b$  then equation (25) reduces to

$$z = \frac{2P_0 a}{c \sigma} \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^2 \{J_1(\lambda_i)\}^2} \int_0^t \psi(\eta) \sin \frac{c \lambda_i}{a}(t-\eta) d\eta,$$

where now the sum extends to all the positive roots of the equation

$$J_0(\lambda_i) = 0. \quad (26)$$

Similarly if  $b \rightarrow 0$ , and  $P_0 \rightarrow \infty$  in such a way that the total load  $\pi b^2 P_0$  tends to a finite limit  $F$ , we have

$$z = \frac{F a}{\pi c \sigma} \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^2 [J_1(\lambda_i)]^2} \int_0^t \psi(\eta) \sin \frac{c \lambda_i}{a}(t-\eta) d\eta, \quad (27)$$

the sum again taken over the positive roots of equation (26).

For example, if we take  $\psi(t)=1$  for  $t \geq 0$  then in the general case  $b \leq a$ ,

$$z = \frac{2P_0 a b}{T} \sum_{\lambda_i} \frac{J_1(b \lambda_i/a) J_0(r \lambda_i/a)}{\lambda_i^3 \{J_1(\lambda_i)\}^2} \left(1 - \cos \frac{c t \lambda_i}{a}\right),$$

so that if  $b=a$

$$z = \frac{P_0}{T} \left\{ 2a^2 \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^3 J_1(\lambda_i)} - 2a^2 \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^3 J_1(\lambda_i)} \cos \frac{c t \lambda_i}{a} \right\}.$$

Now using the result (11) in formula (4) we have

$$2a^2 \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^3 J_1(\lambda_i)} = \frac{1}{4}(a^2 - r^2),$$

giving, finally,

$$z = \frac{P_0}{T} \left\{ \frac{1}{4}(a^2 - r^2) - 2a^2 \sum_i \frac{J_0(r \lambda_i/a)}{\lambda_i^3 J_1(\lambda_i)} \cos \frac{c t \lambda_i}{a} \right\}$$

taken over the positive roots of equation (26), in agreement with the result given by Carslaw and Jaeger.

Similarly, if  $\psi(t) = \sin \omega t$  the appropriate solution of equation (22) is

$$\bar{z}_j = -\frac{P_0}{\sigma} \cdot \frac{b J_1(b \xi_i)}{c \xi_i^2} \cdot \frac{\omega \sin c \xi_i t - c \xi_i \sin \omega t}{c^2 \xi_i^2 - \omega^2},$$

whence

$$z = \frac{P_0 c^2}{T \omega^2} \sin \omega t \left\{ -\frac{2b}{a^2} \sum_i \frac{1}{\xi_i (1 - c^2 \xi_i^2 / \omega^2)} \cdot \frac{J_0(r \xi_i) J_1(b \xi_i)}{[J_1(a \xi_i)]^2} \right\} \\ + \frac{2P_0 \omega c}{T} \cdot \frac{b}{a^2} \sum_i \frac{\sin c \xi_i t J_0(r \xi_i) J_1(b \xi_i)}{\xi_i^2 (\omega^2 - c^2 \xi_i^2) [J_1(a \xi_i)]^2} \quad (28)$$

Substituting from equation (12) into the inversion formula (4) we obtain the expansion

$$-\frac{2}{a} \sum_i \frac{J_0(r \xi_i)}{\xi_i (1 - c^2 \xi_i^2 / \omega^2) J_1(a \xi_i)} = \frac{J_0(r \omega / c)}{J_0(a \omega / c)} - 1$$

so that if  $b=a$  equation (28) reduces to

$$z = \frac{P_0 c^2}{T \omega^2} \sin \omega t \left\{ \frac{J_0(r \omega / c)}{J_0(a \omega / c)} - 1 \right\} + \frac{2P_0 \omega c}{T a} \sum_i \frac{\sin c \xi_i t \cdot J_0(r \xi_i)}{\xi_i^2 (\omega^2 - c^2 \xi_i^2) J_1(a \xi_i)},$$

the summation extending to all the positive roots of equation (3).

### § 5. A Problem in the Conduction of Heat.

To illustrate the use of the second inversion formula (6) we shall apply it to the solution of a problem arising in the theory of the conduction of heat in a circular cylinder of solid section. The cylinder, which is of radius  $a$ , has an initial temperature  $f(r)$ ; radiation takes place at the surface  $r=a$  into a medium maintained at zero temperature. The temperature at time  $t$  is determined by the system of equations

$$\frac{\partial \theta}{\partial t} = \kappa \left( \frac{\partial^2 \theta}{\partial r^2} + \frac{1}{r} \frac{\partial \theta}{\partial r} \right) \quad 0 \leq r < a, \quad t > 0, \quad (29)$$

$$\theta = f(r), \quad 0 \leq r < a \text{ at the instant } t=0, \quad (31)$$

$$\frac{\partial \theta}{\partial r} + h \theta = 0, \quad \text{when } r=a, \quad \text{for } t > 0 \quad (32)$$

Multiplying equation (29) throughout by  $r J_0(\xi_i r)$ , choosing  $\xi_i$  to be a root of equation (5) with  $\mu=0$ , and inserting the boundary condition (32) into equation (17) we have

$$\frac{1}{\kappa} \frac{d \bar{\theta}_J}{dt} + \xi_i^2 \bar{\theta}_J = 0, \quad (33)$$

whence we obtain, by means of equation (31),

$$\bar{\theta}_J = e^{-\kappa \xi_i^2 t} \int_0^a \eta f(\eta) J_0(\xi_i \eta) d\eta. \quad (34)$$

Inverting equation (34) by means of the formula (6) we have, finally,

$$\theta = \frac{2}{a^2} \sum_i \frac{\xi_i^2 e^{-\kappa \xi_i^2 t}}{h^2 + \xi_i^2} \cdot \frac{J_0(\xi_i r)}{\{J_0(\xi_i a)\}^2} \int_0^a \eta f(\eta) J_0(\xi_i \eta) d\eta,$$

the summation extending over all the positive zeros of the function

$$\xi_i J'_0(\xi_i a) + h J_0(\xi_i a).$$



§ 6. *The Motion of Viscous Fluid in an infinite Cylinder.*

We shall conclude by giving an illustration of the use of the second type of finite Hankel transform, defined by equation (7). Suppose viscous fluid is contained between two infinite co-axial circular cylinders of radii  $a$  and  $b$ ; the fluid is set in motion by the outer cylinder  $r=b$  starting to rotate with uniform angular velocity  $\Omega$  at the instant  $t=0$ , the inner cylinder being kept at rest. If we denote the velocity of the fluid by  $v$  then  $v=\Omega b$  when  $r=b$  and  $v=0$  when  $r=a$ , and  $v$  at any other radius is determined by the differential equation

$$\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} - \frac{1}{r^2}\right)v = \frac{1}{\nu} \frac{\partial v}{\partial t}, \quad a < r < b, \quad t > 0, \quad \dots \quad (35)$$

where  $\nu$  denotes the kinematical viscosity.

Multiplying both sides of equation (35) by  $J_1(\xi_i r)G_1(\xi_i a) - G_1(\xi_i r)J_1(\xi_i a)$  and integrating with regard to  $r$  from  $a$  to  $b$  we have, on putting  $\mu=1$ ,  $f(a)=0$ ,  $f(b)=\Omega b$  into equation (18),

$$\frac{1}{\nu \xi_i^2} \cdot \frac{\partial \bar{v}_H}{\partial t} + \bar{v}_H + \frac{\Omega b}{\xi_i^2} \cdot \frac{J_1(\xi_i a)}{J_1(\xi_i b)} = 0. \quad \dots \quad (36)$$

Now  $\bar{v}_H=0$  when  $t=0$  so that the appropriate solution of (36) is

$$\bar{v}_H = -\frac{\Omega b}{\xi_i^2} \cdot \frac{J_1(\xi_i a)}{J_1(\xi_i b)} (1 - e^{-\nu \xi_i^2 t}).$$

Hence from equation (8)

$$v = -\sum_i 2\Omega b \cdot \frac{J_1(\xi_i a)J_1(\xi_i b)}{J_1^2(\xi_i a) - J_1^2(\xi_i b)} (1 - e^{-\nu \xi_i^2 t}) \{J_1(\xi_i r)G_1(\xi_i a) - G_1(\xi_i r)J_1(\xi_i a)\}.$$

Now from equations (19) and (18)

$$\frac{\Omega b^2}{r} \cdot \frac{r^2 - a^2}{b^2 - a^2} = -2\Omega b \sum_i \frac{J_1(\xi_i a)J_1(\xi_i b)}{J_1^2(\xi_i a) - J_1^2(\xi_i b)} \{J_1(\xi_i r)G_1(\xi_i a) - G_1(\xi_i r)J_1(\xi_i a)\},$$

so that

$$v = \frac{\Omega b^2}{r} \cdot \frac{r^2 - a^2}{b^2 - a^2} + 2\Omega b \sum_i \frac{J_1(\xi_i a)J_1(\xi_i b)}{J_1^2(\xi_i a) - J_1^2(\xi_i b)} \{J_1(\xi_i r)G_1(\xi_i a) - G_1(\xi_i r)J_1(\xi_i a)\},$$

where the sum extends over all the positive roots of the equation

$$J_1(\xi_i b)G_1(\xi_i a) - G_1(\xi_i b)J_1(\xi_i a) = 0.$$

I am indebted to the Director General of Scientific Research and Development, Ministry of Supply, for permission to publish this note.

## References.

- (1) Doetsch, G., *Math. Annalen*. cxii. pp. 52-68 (1935).
- (2) Kneitz, H., *Math. Zeits.* xlv. pp. 266-291 (1938).
- (3) Brown, H. K., *Bull. Amer. Math. Soc.* xlviii. p. 522 (1942); *Journ. App. Physics*, xiv. p. 609 (1943).
- (4) Sneddon, I. N., *Proc. Camb. Phil. Soc.* xli. p. 29 (1945).

- (5) Carslaw, H. S., and Jaeger, J. C., 'Operational Methods in Applied Mathematics.' (Oxford, 1941.)
- (6) Watson, G. N., 'A Treatise on the Theory of Bessel Functions,' Chapt. xviii. (Cambridge, 1922.)
- (7) Carslaw, H. S., and Jaeger, J. C., *op. cit.* pp. 160-161.

The Department of Natural Philosophy,  
The University of Glasgow.

#### IV. *The Measurement of Pressure on the Sea Bed.*

By E. G. RICHARDSON\*.

[Received March 20, 1946.]

[Plate I.]

##### *Introduction.*

ALTHOUGH the measurement of hydrostatic pressures in the laboratory raises no difficulties, to measure the underwater pressure exerted on the sea bed is not so simple, firstly because, owing to tide and waves the pressure is varying and, secondly, because differences between the pressure inside the instrument and the sea outside may give rise to leaks which are difficult of access for repair.

On both accounts it is advisable that the mean pressure inside the instrument shall be equal to the pressure in the sea outside. Then, the instrument actually records deviations from the mean and is working at maximum sensitivity.

The results of such pressure measurements provide an interesting check on the theory of surface waves and, with this theory established, enable the instrument to be used as a wave height recorder.

##### *Theory of Measurement.*

Suppose at the surface a system of waves periodic in  $x$  and of amplitude  $a$ , wave-length  $2\pi/k$ , pulsance  $\omega$ . If the depth to the bed is  $h$  the velocity potential of the motion at a point  $y$  (measured from the surface) is

$$\phi = \frac{ga}{\omega} \frac{\cosh k(y+h)}{\cosh kh} \cdot \cos(kx - \omega t) \dagger.$$

Applying Bernoulli's principle, but neglecting the square of the current,

$$\frac{p}{\rho} = \frac{\partial \phi}{\partial t} - gy,$$

\* Communicated by the Author.

† Cf. Lamb, 'Hydrodynamics,' § 228.

where  $p$  is the pressure at depth  $y$  in a liquid of density  $\rho$ ; so that the pressure amplitude

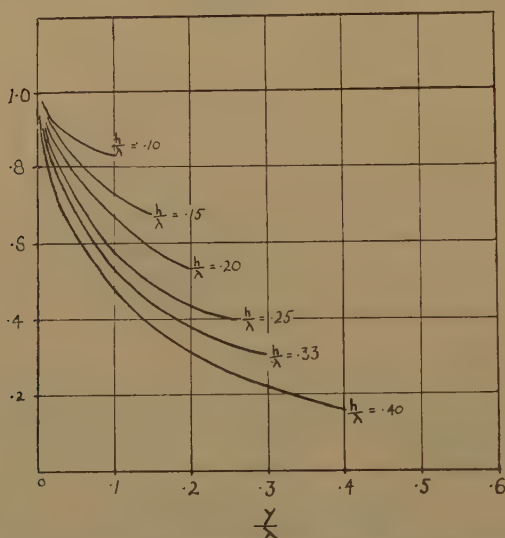
$$\Delta p = \rho g a \frac{\cosh k(y+h)}{\cosh kh} \\ = \rho g a (\cosh ky + \sinh ky \tanh kh) \quad . \quad . \quad . \quad (1)$$

At the bed

$$\Delta p_0 = \rho g a \operatorname{sech} kh \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The quantity  $(\cosh ky + \sinh ky \tanh kh)$  is plotted for various values of  $\frac{h}{\lambda}$  against  $\frac{y}{\lambda}$  ( $y$  negative) on fig. 1. Short waves record a small pressure

Fig. 1.



Variation of pressure amplitude with depth and wave-length.

(for a given amplitude) at the bed, for when  $h > .8\lambda$ ,  $\tanh kh = 1$ , approx. and negligible hydrostatic pressure amplitude is felt at the bed. In a "mixed" sea then the swell is felt at full effect but the little waves pass ignored at the bottom.

### Theory of the Manometer.

As it was the pressure of waves rather than of tidal motion which it was desired to record, an instrument had to be designed which would reach equilibrium, as to its internal pressure, with changes in hydrostatic depth due to the tides, but record—with practically negligible loss—the pressure due to surface waves.

To accomplish this the principle of the equalisation chamber and pressure leak was used. Fig. 2 shows this applied to two types. The

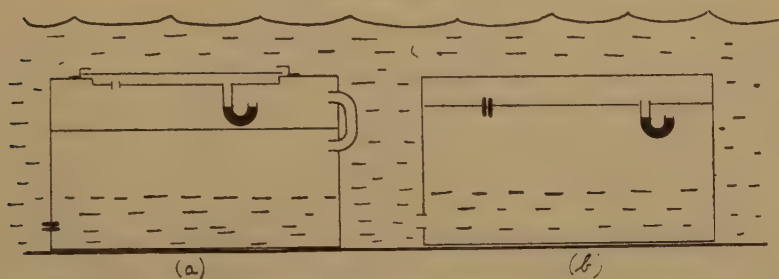


equalisation pot is a large chamber into which the sea can leak slowly at the bottom. At the top of this an air pipe conducts to the measuring chamber which in the type A has a rubber diaphragm over a metal dish at the top, connected via a capillary tube or bleed hole into the main chamber, which contains air at the prevailing mean hydrostatic pressure. The pressure gauge measures the air-pressure difference between the capsule and the main chamber. When the apparatus is lowered into the sea, water leaks into the equalisation chamber until the mean pressure throughout the inside is equal to that outside.

When a sudden excess or defect of pressure is applied to the diaphragm, pressure builds up in the capsule, because air cannot leak away fast enough through the capillary. The manometer then records the difference between the pressure in the capsule and that in the equalisation chamber.

The rubber diaphragm is not necessary if the sea can pass freely to and from the equalisation chamber, so as to apply the pressure freely to the air in the larger chamber. The walls of the chamber above the capsule

Fig. 2.



Types of pressure recorder.

are then made rigid (fig. 2(b)). Thus in the first type, entry of water to the equalisation chamber is restricted by a slow leak and the extra pressure applied through the diaphragm, whereas in the second type unrestricted entry of water ensures that the sea pressure is applied directly to the large chamber.

The recorder may be considered as a large volume of air  $V$  in the main chamber and a small volume  $v$  in the capsule connected by a tube of radius  $r$  and length  $l$  containing air of viscosity  $\eta$  and density  $\rho$ . The impedance of the system to alternating fluctuations of pressure at pulsantance  $\omega$  applied to either chamber is, neglecting the inertance of the air in the "leak,"

$$\sqrt{\left[\left(\frac{8l\eta}{\pi r^4}\right)^2 + \left(\frac{\partial p}{\partial \rho} \cdot \frac{\rho}{\omega}\right)^2 \left(\frac{1}{V} + \frac{1}{v}\right)^2\right]}$$

The amplitude of the pressure difference between the two chambers as recorded by the gauge, per unit amplitude of pressure in the sea outside, is the reciprocal of

$$\sqrt{\left[1 + \left\{\frac{\partial p}{\partial v} \cdot \frac{\pi r^4}{8vl\omega} \left(\frac{1}{V} + \frac{1}{v}\right)\right\}^2\right]} \dots \dots \dots (3)$$

when  $\nu$  is written for the kinematic viscosity of air. Here we neglect the slight interchange of volume between the chambers as the gauge "responds" and in the case of type A, of the change in  $\nu$ , due to the motion of the rubber diaphragm. We also assume that the latter has no motional impedance, i. e., is perfectly flexible.

Two important characteristics of the recorder can be deduced from (3); firstly, in order to record motions of low frequency adequately, the resistance of the leak, which is proportional to  $l/r^4$  must be large, but not so large that the instrument does not reach equilibrium under the very slow periods of the tides; secondly, that, due to reduction in the volume of the chamber and the increase in kinematic viscosity of the air with pressure, the sensitivity of the recorder decreases with increasing depth of bed. Further, at any depth the long waves will be less adequately recorded than the short ones (owing to the less value of  $\omega$  for the former). This to a certain extent offsets the difference, in the opposite sense, between the theoretical pressure amplitudes exerted by long and short waves (cf. [2]).

### *Applications of the Apparatus.*

The recording of "submarine weather" is not the only application of a pressure recording instrument under the sea.

Forty years ago Gaillard\* fitted a rubber diaphragm unit to the pile of a jetty where the Duluth canal debouches into Lake Superior. He was concerned about the forces to which underwater structures would be subjected by waves during and after a storm. A water column connected the diaphragm to a Bourdon gauge above (on the stethoscope principle). Gaillard was not troubled with tidal ebb and flow, and although his gauge was insensitive and the water column in the pipe leading to the gauge had considerable inertia, he reckoned that his measurements were correct for wave periods greater than one second. Nevertheless his recorded pressures were mostly lower than the theoretical pressures.

It is obvious that, once the validity of formulæ (1) and (2) had been established, an instrument such as that described in this paper might be used, after applying the appropriate frequency and capacity corrections, to record the wave height on the surface above. This application has some importance in connection with the navigation of hydroplane boats and the take-off and landing of seaplanes.

Pabst† so used a D.V.L. scratch recorder—an instrument in which the pressure on a diaphragm causes a record to be scratched by a diamond on a slowly moving steel plate—in a vessel suspended by a wire from a buoy riding on the surface, to measure wave height and length.

To the record provided by the instrument, corrections must be applied, in the light of the graphs of fig. 1, to get the wave height at the surface.

---

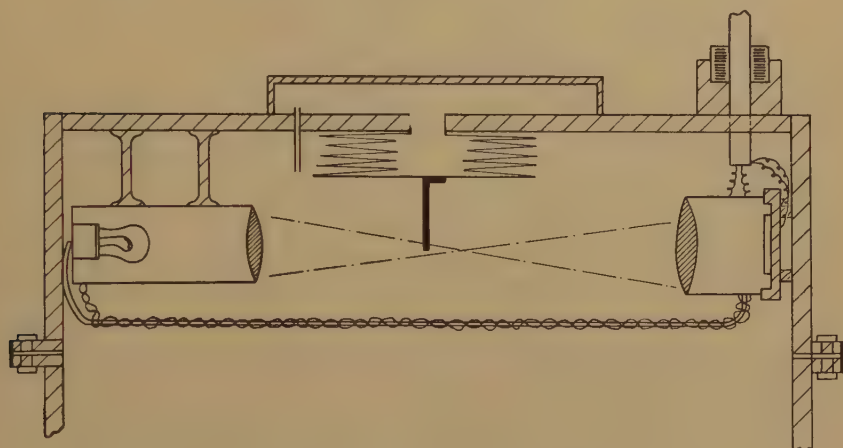
\* Professional Paper No. 31, Corps of Engineers of U.S. Army, 1904.

† *Zeits. f. Flugtechnik*, xxiv. p. 598 (1933).

*Description of Pressure Recorder.*

Of the two designs described (fig. 2), A is probably the more sensitive, in spite of the disadvantage of the rubber diaphragm exposed to the sea, and was adopted for the first measurements which were carried out in 1941 while the author was a temporary member of the Scientific Pool of the British Admiralty. In order to get good response to small differences of pressure between the two chambers a sylphon bellows made by Messrs. Negretti and Zambra was used as pressure gauge, and the motion of a tag attached to the lower end of the bellows recorded photo-electrically. An apparatus involving a rubber diaphragm proved to be liable to develop leaks or to be punctured and so unsuitable for immersion for prolonged times in the sea, and in a later form (1944), design B was followed.

Fig. 3.



Details of second form of recorder.

Fig. 3 is a diagram of the later apparatus, of which a photograph showing the interior is given on the Plate, the large equalisation chamber below the recorder being omitted from the sketch. The latter has a volume of 60 litres, that over the sylphon 2 litres, while the "leak" is a piece of copper tubing 0.5 mm. diameter and 4 cm. long of a type used in rate-of-climb meters on aircraft. The bellows is screwed into a central hole beneath the capsule, and the light metal tag soldered to the lower end intercepts a focused beam of light which afterwards falls on a selenium-rectifier photo-electric cell. A submarine cable brings in leads for lamp and cell. The whole apparatus is mounted on a heavy steel sinker with space below for the ingress of water to the large chamber and shackles for lowering into the sea.

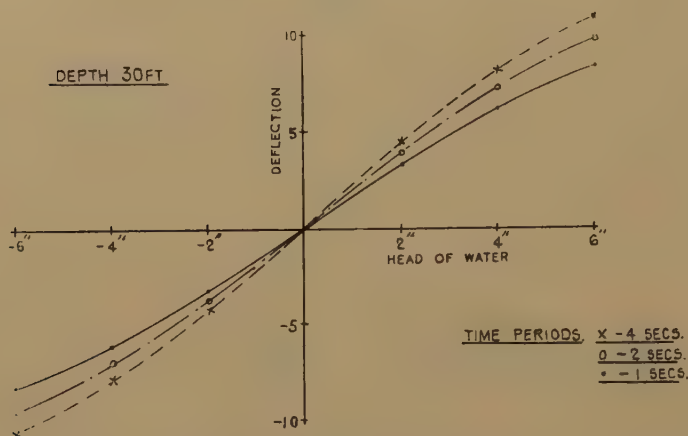
*Calibration of Recorder.*

The first test was to find, by blowing up or partially exhausting the capsule through the leak, that there was a linear relationship between



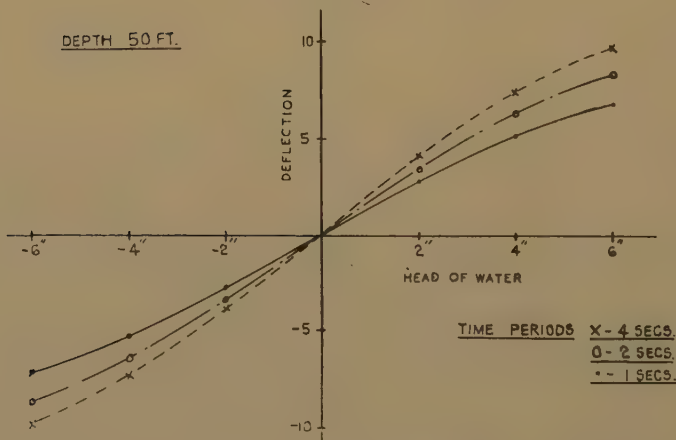
steady pressure differences and the motion of the bellows as recorded by a sensitive galvanometer connected to the photo-electric cell. This was found to be so except when the bellows reached the stage of maximum inflation or deflation—regions which were not expected to be over-run in practice.

Fig. 4.



Response curves of recorder.

Fig. 5.



Response curves of recorder.

The formula (3) shows that the response varies with both  $V$  and  $\omega$ . This was checked by rhythmic lowering and raising of the apparatus on a winch, about a mean position in a fire-tank 50 ft. high or, for shallower immersions, by keeping the apparatus still and rhythmically raising and lowering the surface of the water by a large bucket dipped beneath it.

The graphs (figs. 4, 5) exhibit the relation between galvanometer excursions and pressure amplitudes at various frequencies and two hydrostatic pressures. These fit a formula of type (3) quite well.

Another method of calibration is afforded by pumping up the capsule and bellows to a steady pressure and suddenly releasing it, recording the movement of the bellows as it deflates. The "damping coefficient" of the system, which can be derived from such a record, is the reciprocal of the product of the resistance component (of the leak) and the capacitance (of the reservoir) as in the analogous electrical case of the discharge of a condenser. Knowing the "time-constant" of the system in this form, records at various frequencies of simple wave-forms can be "corrected" for lag by means of it.

Tests of the validity of (2) were made, by courtesy of the Superintendent, in the "old" Haslar Experimental Ship Tank, in which model waves of known height and length can be set up.

### *Tests of the Apparatus in the Sea.*

The early form of the apparatus was tested in 1941 in the sea off the end of Ventnor Pier (Isle of Wight) and off Southsea Pier in about 6 fathoms (at high tide). The former corresponds to a site exposed to the up-channel swell from the Atlantic Ocean, the latter to a sheltered site in the Solent, protected by the Isle of Wight from Channel swell. Leads were taken to the battery and to the galvanometer in a dark room at the pier head. It was arranged that deflections of the galvanometer should be recorded on a slowly moving paper camera.

Typical records are shown on Pl. I. Two paper speeds are involved, and these are given at the sides of the photographs. The sensitivity at  $\omega = \infty$  is 1 cm. of vertical deflection for  $\frac{1}{4}$  in. of water. The records involve waves up to 40 ft. long and 4 ft. in height. Looking at the surface, one noticed waves much shorter and shallower than these, whose pressure amplitudes at the beds would be too small to show on the records. The records are in fact indicative of what is usually known as "swell," rather than of the superficial wavelets more apparent to the uncritical observer's eye.

The velocity ( $c$ ) of surface waves is a function of the wave-length, being given by the formula

$$c^2 = \left( \frac{g}{k} + \frac{\sigma k}{\rho} \right) \tanh kh, \quad . . . . . (4)$$

where  $\sigma$  is the surface tension and  $k = \frac{2\pi}{\lambda}$ , as before. In consequence

of (4), in a mixed wave-system the records show clearly a "group velocity" or interference pattern in the pressure amplitude: time record which can be ascribed to the overrunning of the shorter waves by the longer. Thus records 1 and 2 show each two wave-systems of group 1.5 and 0.75 mins. respectively. This type was noticeable in calm

weather. During and after a storm the waves became incoherent (record 3).

A knowledge of the type of "swell" to be observed under various topographical and climatic conditions round the coasts is obviously of importance to hydrographers and work on these lines was and is being continued, after the author left the Admiralty service, by a scientific team at the Admiralty Research Laboratory\*.

#### *Pressure Records of Ships.*

Inasmuch as a ship in motion displaces water and sets up a wave-system, of which the bow wave and the stern wave are the most prominent, one would expect a moving ship to produce a disturbance of the prevailing fluctuations in pressure on the sea bed set up by the wind-produced wave system. Such effects were, in fact, observed at Southsea, where the recorder happened to lie in the fairway for ships passing in and out of Portsmouth Harbour. Pl. I. shows two such records, one (No. 5) caused by a trawler passing at a speed of about 6 knots about ten yards to one side, and another (No. 4) of a motor vessel passing overhead at 18 knots. (The vertical strokes on the records indicate the nearest instant of approach of the bow to the instrument.) These are typical; there is first a small positive pressure accompanying the bow wave, then a deep trough of suction and another positive peak associated with the stern wave; finally, damped oscillations following in the wake. The bed record is not an exact copy of the wave-system as seen alongside the vessel, since the latter's effect at the bed is complicated by the considerable cross motion imparted to the water by the displacement of the hull; the waves due to a ship are not, in fact, plane but have V-shaped fronts.

It may be remarked in passing that these pressures at the sea bed are, of course, accompanied by water currents, related to the pressure by Bernoulli's principle. On one occasion a hot-wire current meter, connected to its own galvanometer, was lowered on the same sinker with the pressure recorder and gave a water movement record which was a faithful companion of the pressure record, regarded from the Bernoulli aspect.

#### *Summary.*

An account is given of the construction of an apparatus for measuring the pressure produced by the sea on its bed, based on the capacity and leak system. The calibration of the instrument follows from classical hydrodynamic theory, which is put to practical test. The use of the apparatus for wave-height recording is illustrated by some measurements made in the English Channel and its possible use to record "surface weather" discussed.

---

\* Cf. 'Nature,' clvii. p. 165 (1946).



V. *Colour Vision.*

By JOHN H. SHAXBY, D.Sc., F.Inst.P., Physiology Institute,  
University College, Cardiff \*.

[Received June 15, 1945.]

1. VARIOUS theories of colour vision have been propounded, of which the Young-Helmholtz Trichromatic, with some later additions and modifications, is to-day the most widely held. In company with others, this theory is concerned with the retinal mechanisms which render possible the differences in visual sensations which we term differences of colour. The further question, as to how such retinal differentiations are carried through to the visual cortex, in which these differences take their final sensory form, has received much less attention. It is clear that retinal activities in themselves cannot tell the whole story; the nervous messages generated in the retina must retain their individuality in their passage along the fibres of the optic nerve to their final central destination.

The effective retinal elements are beyond dispute the cones, the only receptors found in the fovea, and the preponderating units in the immediately surrounding parafoveal area; as we move outward from the fovea to the almost colour-blind extreme periphery of the retina, the gradual weakening of the colour sense runs parallel with the lessening of the cone population.

The Young-Helmholtz theory may be expressed in terms of three species of cone, responsible respectively for the red, green and blue sensations, or in terms of three types of photo-chemical activity in each cone; other possible variations are clearly possible, for instance red and green mechanisms in one kind of cone, blue in another, but the evidence for any particular alternative is too slender to be decisive. Again, Koenig (1894) observed the fact that with small areas of stimulation of the fovea, the latter is dichromatic. Recently Willmer (1944) independently observed this dichromasy, and Willmer and Wright (1945) and L. C. Thomson and Wright (in a recent paper to the Colour Group of the Physical Society) have given quantitative analysis of this tritanopia of small areas at and near the centre of the fovea. Hartridge (1945) finds that this dichromasy is not confined to the fovea, but is a general feature of the retina with sufficiently small sources of light. But these very interesting and suggestive observations do not affect the general position that some sort of retinal triplicity best describes the performance of the eye in matters of colour.

---

\* Communicated by the Author,

The trichromatic theory is accepted as giving a very satisfactory account of the experimental facts of normal colour-matching and colour-mixture. Its inadequacy in explaining such phenomena as simultaneous contrast and after-images has probably been overstressed; our growing knowledge of the effects of adaptation to particular wave-lengths, and especially of the interactions of retinal structures, notably in the lateral associations permitted in synaptic layers, etc., gives opportunity for explanation in terms of summation and particularly inhibition, familiar in other neural relationships.

On the other hand, the completeness with which the theory accounts for the facts of colour-matching and -mixture is not in itself conclusive evidence of its truth; the theory is itself a restatement in physical terms of the experimental data, in such colour equations as

$$(\alpha + \beta + \gamma)(X) = \alpha(R) + \beta(G) + \gamma(B),$$

the algebraic expression of the facts that light of a colour designated as X can be exactly matched by a suitable mixture of three primaries, R, G, and B, of brightnesses which add up to that of X, and in proportions defined by the ratios of  $\alpha$ ,  $\beta$ , and  $\gamma$ . The equation also conveys the fact that in general three such primaries are necessary, and that no range of mixtures of two only can match the full series of colours. But this is not a proof that the retinal mechanism is three-fold in nature. It may be possible that the three terms on the right-hand side of the equation are necessary to fulfil the two required conditions, equality of brightness and identity of colour. That theoretically this may be the case has been shown in an earlier paper in this magazine (Shaxby, 1943), where the hypothesis was put forward that an exact match demands not only equality of the action-current energy (ensuring a brightness match), but also an equality of the make-up of the individual action-current units resulting from the matched illuminations.

These units are necessarily of equal magnitudes, as stated in the All-or-None Law, applicable to action-currents in general, but their equal energies, built up of a number of electronic charges as units, may rest on a larger number of such units of smaller individual energy, or on a smaller number of greater individual energy. Thus blue light of a given energy results from quanta of (visually) maximal amount, so that the electrons corresponding to blue light can build up the required constant unit of action-current by a smaller number than can red light with its smaller quanta. A perfect colour match occurs when the electronic make-up of the unit is the same for both fields of illumination. For a match between a spectral monochromatic light of wave-length  $\lambda$  and a mixture of the three spectral primaries of wave-lengths  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , the necessary expressions were shown to be the ordinary equation for equality of brightness,

$$S_\lambda = S_1 + S_2 + S_3, \quad \dots \dots \dots (1)$$

where  $S_\lambda$  denotes the brightness of the given monochromatic light,  $S_1$ ,

$S_2$  and  $S_3$  those of the three matching primaries, and a second, the true colour equation, expressing the match in colour.

From the condition of identity of electronic composition of the discharge units of action currents arising from the matched lights, this equation was found to be

$$S_1 \cdot \lambda(1-k\lambda) = S_1 \cdot \lambda_1(1-k\lambda_1) + S_2 \cdot \lambda_2(1-k\lambda_2) + S_3 \cdot \lambda_3(1-k\lambda_3),$$

where  $k$  is an "efficiency" factor for the ejection of an electron from a molecule of the photo-sensitive substance of the retina under the impact of a quantum. The equation was shown to give a satisfactory description of the matches of 25 spectral colours throughout the visible spectrum, as determined by W. D. Wright.

In view of the quantal nature of the processes concerned, it seems preferable to replace the efficiency factor  $(1-k\lambda)$ , etc., by a term expressing the probability of escape of an electron due to retinal bombardment by light quanta. The colour equation is thus modified (see below) to the form

$$S_1 \cdot \lambda e^{-K\lambda} = S_1 \cdot \lambda_1 e^{-K\lambda_1} + S_2 \cdot \lambda_2 e^{-K\lambda_2} + S_3 \cdot \lambda_3 e^{-K\lambda_3}, \quad (2)$$

where  $K$  is the factor defining the probability of escape. In this form the equation is found to be in as good or slightly better accord with experimental data as the first; it also satisfies the data for a series of cases of trichromatic anomaly.

The colour equation may for brevity be written as

$$Q_\lambda = Q_1 + Q_2 + Q_3,$$

where the several terms represent the quantities  $S_1 \cdot \lambda e^{-K\lambda}$ , etc. These  $Q$ 's may conveniently be termed the Colour Quotients, since they represent the numbers of effective electrons of the matched light and the three matching primaries, obtained by dividing the sensory energies (luminosities) by the magnitude of the relevant quanta.

It is to be noted that while the hypothesis demands no retinal triplicity, it in no way excludes it. The three primary terms in the equations of a complete colour match are necessary because two equations have to be satisfied, and this cannot be done if only two primaries are available. For if we omit the third term on the right in equations (1) and (2) above, we see that, combining the two,  $(S_1 + S_2)\lambda e^{-K\lambda} = aS_1 + bS_2$ , writing  $a$  for  $\lambda_1 e^{-K\lambda_1}$  and  $b$  for the similar function of the second primary. For any given  $\lambda$  this equation can only hold for one particular value of the total brightness. But if we add a third primary the corresponding equation yields unique values of  $S_1$ ,  $S_2$  and  $S_3$  to satisfy both (1) and (2), and these values are in agreement with the experimental data of spectral colour-matches.

2. As has been remarked, the trichromatic theory makes no attempt to explain how retinal triplicity is preserved through the nerve fibres to the cortex, as it must be if the differentiation we know as colour difference is to appear in sensation. The hypothesis outlined above suggests a



way in which this differentiation may be maintained. In fact, since we cannot explain colour differences by cortical topography, which is concerned with local signature, nor by variations in the rhythm of action current discharges, which are concerned with brightness differences, it seems necessary to look for some such differences, within the actual units of action-current discharge, as have been postulated above.

But there is an increasing amount of support for a polychromatic view of retinal activity. The researches of Hartline and of Ragnar Granit seem to give definite evidence of differentiation of cones in animal retinae, perhaps of a number of types rather than strictly three. More recent accounts, in a letter to Prof. Hartridge, in an article in 'Nature,' and in the Guthrie Oration of the Physical Society for 1945, confirm this view. Granit reports the finding of several types of receptor, at wavelengths ( $\mu$ ) 0.60 (common), 0.58 (less common), 0.54, 0.52, 0.50, 0.46 and 0.33. A further puzzling feature of Granit's work is the inconstancy of individual cones, which appear to change in type according to the state of light adaptation of the eye.

But perhaps the strongest support for the trichromatic theory comes from its power to account for cases of abnormal colour vision, those of Dichromats and Trichromatic Anomalous subjects. Recent work by F. H. G. Pitt (1944) gives a striking instance of this; he shows that the colours ranked as matches by deuteranopes are those which the trichromatic theory would predict on the supposition that in these cases there has been a failure of differentiation of the longer wave receptors into the two classes for red and green, a hypothesis which has been held on general grounds by other workers. Similarly, the protanope's series of matching colours is shown to be that which would result from absence of the specialized red receptors of the trichromatic theory.

In a later paper (Pitt, 1945) it is shown that a particular class of monochromats, those possessing foveal cones, have visibility curves, etc., which accord closely with those to be expected if they are both protanopic and tritanopic.

The power of the theory thus to predict unequivocally the various observed features of abnormal colour vision seems strongly in its favour. If then we adopt it, as applied to the retina, and add to it the hypothesis here advanced of the constitution of the individual units of action-current in the optic nerve fibres, we can formulate a view of colour vision which includes at any rate a great part of the phenomena of the appreciation, discrimination and matching of colour.

It may be first noted that the evolution of the sense of colour differences through the varying electronic content of action-current units would be aided in its inception by a retinal differentiation of three types of cone (or of photo-chemical substance), or even of two such types, reacting respectively to the longer and the shorter waves of the visible spectrum. Such a primitive division into Blue and Yellow receptors was suggested by Mrs. Ladd-Franklin (1929), the longer wave type being supposed to split later into two, the specifically Red and Green receptors,

If there were such groups it must of course follow that the action-current units would tend to a parallel demarcation in respect of the number of electrons constituting them, with corresponding differences in their stimulation of cortical cells, and of differences in the resulting sensations. For change of the stimulating wave-length would clearly result in a greater change in the number of constituent electrons of the fixed action-current packet than would occur if the retinal receptors were unspecialized. Once this qualitative, or rather numerical, difference had been "recognized" on the plane of sensation, further improvement in the power to differentiate between smaller and smaller differences of electronic content would become a matter of "physiological education," as in the case of other human faculties of observation and differentiation.

The splitting of the longer-wave receptors into two sub-groups would accentuate the process, and in particular would tend to a high power of discrimination in the region of the spectrum where two types of receptor were more or less equally represented in the make-up of a mixture. Thus the existence of distinct Red and Green elements would result in especially fine discrimination in the yellow, since there would here be a maximum turnover in the relative proportions of electrons derived from the Red and Green receptors; a parallel large change in the number of constituent electrons, per unit action-current packet would occur. Thus the high discrimination near the sodium D-lines, initiated by the great change with wave-length of the relative numbers of stimulated Red and Green receptors, will be maintained to the sensory stage by the correspondingly large change in the number of electrons associated with the constant energy packet of action-current.

3. Summing up, we see that when light falls upon the retina the receptor units react photo-chemically to convert the energy of the quanta into the electronic energy of action-currents in nerve fibres by the process of ejection of electrons, the bricks building up the all-or-none unit of the nerve discharge. If the exciting wave-length changes, the change in magnitude of the quantum produces a correlated change in the individual electronic energies, and therefore in the number of electrons required to build up the fixed unit of discharge. This change is accentuated at wave lengths where the change-over from one type of cone to another is most marked, *viz.*, in the yellow and in the blue-green, so that the eye comes to possess its greatest powers of colour discrimination for these colours.

If two illuminations excite equal action-currents, that is to say equality of time-interval between successive individual discharges, they result in sensations quantitatively equal; the two fields match in brightness.

But if the wave-lengths concerned are not the same, these equal packets of energy, even though their rhythms of discharge are the same, yet are made up of electrons of differing energies, and therefore contain unequal numbers of them. The resulting difference in cortical stimulation is recognized as a difference in colour. For a colour-match the

distribution of electrons of various energies (or the particular energy in the case of monochromatic lights) must be the same for the two fields. For this reason a spectral colour cannot be matched by any mixture of two other spectral colours, one longer, the other shorter in wave-length. Even though the average energy of the electrons arising from the mixed beam were the same as that from the homogeneous beam, the former has a spread of these electrons of different energies about the mean. Now if electrons of all energies corresponding to the visible spectrum are simultaneously excited in certain proportions, the resulting sensation is that of white light; any spread of energies among the electrons is thus some degree of approach towards whiteness, in other words results in a partial desaturation of the dominant colour. A monochromatic beam must therefore have some proportion of white light added to it, in order that the whole should be capable of giving a colour-match with a mixture of two wave-lengths. Only in this way can the action-currents consist of identical packets. Hence the brightness-match equation of any monochromatic illumination may be written in the form

$$S+W=S_1+S_2,$$

where  $S$  is the brightness of the monochromatic light (let us for definiteness take it as yellow),  $S_1$  that of one of the two appropriate primaries,  $S_2$  that of the other (in this case Red and Green), and  $W$  the brightness of the white light added to  $S$  to make a match possible. Since white light is equivalent to a mixture in definite proportions of the three primaries, let  $W=s_1+s_2+s_3$ . The equation of match thus becomes

$$S \text{ (Yellow)}=(S_1-s_1) \text{ (Red)}+(S_2-s_2) \text{ (Green)}-s_3 \text{ (Blue)}.$$

The equation for any spectral colour is always of this form; three primaries are always required, but one of them must have a negative sign, in other words it has to be added to the spectral colour under examination in order that a mixture of the two other primaries may give a match. The two mixtures then produce action-currents whose individual packets consist of the same number of electrons, of the same energies; they therefore necessarily produce identical cortical stimulations, recognized as a match in colour. (It is not of course suggested that each single nerve fibre must possess this identity of composition in its discharges, but merely that for a reasonably small group there must be equality in the average number of electrons associated with the discharge.)

4. A modification of the colour equation from the form given in an earlier paper has been mentioned above. In the former paper the terms  $(1-k\lambda)$ , etc., were obtained on the supposition that the larger quanta, say of blue light, would be more efficient in releasing an electron than the smaller quanta, say of red light. But in view of the essentially quantum nature of the processes in question, it seems more reasonable to look at the matter from the point of view of the probability of the ejection of an electron, held under molecular restraints (statistically constant), by



the advent of a quantum. The problem is thus similar to that of the ejection of a thermo-ion by the energy of molecular kinetic energy from a hot body, or perhaps of the evaporation of a liquid by the loss from its surface of molecules of kinetic energy greater than some critical value. In any such case the probability of escape takes an exponential form, *e. g.*, in the case of thermions  $e^{-E/kT}$ ;  $E$  is a critical energy, proportional to that to be overcome in escape. In the same way here the appropriate form is  $e^{-E/h\nu}$ , where  $\nu$  is the frequency of the light waves. This may thus be written  $e^{-K\lambda}$ , and the colour equation  $n$  becomes that given (2) above.

The agreement with the observed data of colour vision given by Wright is even slightly better than that of the original form of the equation, tabulated in my former paper, if the value 1.5 is given to  $K$ . It would serve no purpose to repeat Table IV. of that paper with the values derived from (2), merely to show this agreement; the mean error for the whole series of wave-lengths is  $0.45 \pm 0.32$  per cent.

5. Since anomalous trichromats, like normal observers but unlike dichromats, make unique colour matches, it seemed of interest to test the validity of the colour equation (2) for such subjects. The cases studied were those investigated by J. H. Nelson (1938) and F. H. G. Pitt (1935). Nelson's tables give directly the luminosity components needed for the present calculations. Pitt's results are in terms of trichromatic coefficients, and he was kind enough to lend me his original large-scale graphs, from which the luminosities could be calculated with some accuracy, though less for small brightnesses than those from Nelson's data.

Equation (2) was again well satisfied, though of course with values of the equal terms on its two sides differing from those of normal observers for the same wave-lengths. Nelson's subjects were 6 deuteranopes and 1 protanope, Pitt's 3 deuteranopes. Each subject's data satisfy the equation to an accuracy of about 1 per cent. except for the very small values at the blue end of the spectrum, where discrepancies rise to 10 per cent. and at wave-length  $0.41 \mu$  much higher. This is to be expected, from the uncertainties of measurement of such small quantities, and the dependence of the calculated value of the right-hand side of the equation which involves the differences of these measurements. The full tabulation occupies a good deal of space and adds nothing to the arguments advanced above. For the various wave-lengths (25 in all) the colour equation (2) is satisfactorily balanced; the equal values for each wave-length differ systematically from the values given by the average normal observer, recorded in Wright's tables. The differences are considerable, in spite of the fact that the method of observation in Wright's and Pitt's researches necessarily ties together the two sets of observations at the Primary wave-lengths,  $0.65$ ,  $0.53$  and  $0.46 \mu$ , where the quantity  $Q$  is the same for normal and abnormal subjects, *viz.*,  $0.119$ ,  $0.294$  and  $0.0125$  respectively. Deuteranopes differ considerably among themselves in their  $Q$  values, while the protanopic observer shows a

markedly different divergence from the normal. This, of course, is obvious from the original data, and the determination of the colour quotients adds nothing to our knowledge of the nature of deuteranomaly or protanomaly; that equation (2) is satisfied for these subjects who obtain unique though abnormal matches is the significant fact.

#### SUMMARY.

The theory put forward in an earlier paper in this magazine, that the condition of a complete colour-match is that the individual bursts of action-current must contain the same number of electrons for the matched light as for the matching mixture of primaries, is further developed. This view of the nervous activities in colour vision is shown to link well with the trichromatic theory of the retinal processes concerned, in such matters as colour discrimination, the necessity of desaturating a spectral colour in order to obtain a match with a mixture of the three primaries, etc.

The "colour quotients" which are the terms of the equation which expresses the theory, *i. e.*, the measures of the number of electrons associated with an action-current discharge, are somewhat modified from the form given in the earlier paper, in view of the quantum nature of photochemical actions; the new equation fits the facts of colour matching for normal observers slightly better than did the previous form. It is also satisfied for the matches made in cases of Trichromatic Anomaly.

#### *References.*

- Granit, R., 'Nature,' clv. p. 716 (June 16, 1945).  
 Granit, R., Proc. Phys. Soc. lvii. p. 447 (1945).  
 Hartridge, H., 'Nature,' clv. p. 657 (1945).  
 Koenig, A., *Sitzb. Akad. Wiss.* (Berlin), p. 577 (1894).  
 Ladd-Franklin, C., 'Colour and Colour Theories' (1929).  
 Nelson, J. H., Proc. Phys. Soc. l. p. 661 (1938).  
 Pitt, F. H. G., M.R.C. Special Report XIV, Committee on Physiology of Vision, Appendix I. (1935).  
 Pitt, F. H. G., P. R. S. (B) cxxxii. p. 101 (1944).  
 Pitt, F. H. G., 'Nature,' cliv. p. 466 (1945).  
 Shaxby, J. H., Phil. Mag. ser. 7, xxxiv. p. 289 (1943).  
 Willmer, E. N., 'Nature,' cliii. p. 774 (1944).  
 Willmer, E. N., and Wright, W. D., 'Nature,' clvi. p. 119 (1945).

# VI. *A Centrifugal Method of Measuring the Surface Tensions and Interfacial Tensions of Liquids.*

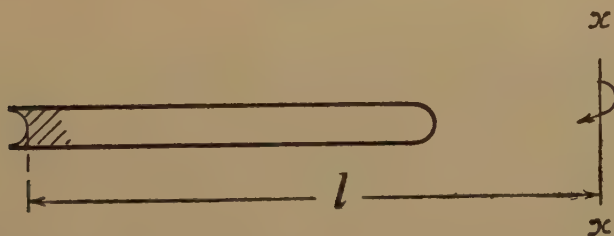
By W. MEYERSTEIN, M.D. (Department of Physiology, University of Birmingham), and J. D. MORGAN, D.Sc.\*

[Received June 25, 1945.]

IN a previous communication we have described a centrifugal method of measuring the surface and interface tensions of liquids †. This method consists in placing radially, on a horizontal table, a tube containing the liquid or liquids, and noting the speed of the table at which movement or liquid occurs. In that communication we described several variants of the method and, in order that the main subject of the present communication may be the more readily understood, we will briefly refer to two of them.

For measuring the surface tension of a liquid a tube, which is open at one end and closed at the other, is filled with the liquid so that a hemispherical meniscus lies at the open end of the tube, as shown in fig. 1.

Fig. 1.



The table is then rotated about its axis  $xx$  at a gradually increasing speed. When the critical speed is reached the liquid in the tube is discharged suddenly and completely. The formula for this case is

$$T = \frac{1}{2} \rho d^2 \pi^2 n^2 l. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Where  $T$  = surface tension.

$\rho$  = density of the liquid.

$d$  = diameter of tube bore in cm.

$n$  = rate of rotation of the table in revs./sec.

$l$  = distance in cm. of the meniscus from the axis of rotation.

For measuring the interfacial tension of two liquids a pair of tubes are used, as shown in fig. 2.

\* Communicated by the Authors.

† Phil. Mag. ser. 7, vol. xxxv. p. 335 (May 1944).



The inner tube is partially filled with the liquid of higher density, and the remaining space is wholly or partially filled with the other liquid (as indicated at *a*). Then the tube is placed in the outer tube, containing a quantity of either of the two liquids. At the critical speed the liquid *a* moves towards the closed end of the inner tube. Provided the meniscus between the two liquids in the inner tube is hemispherical, formula (1) applies also to this case, excepting that  $\rho$  now represents the difference between the densities of the two liquids.

The other variant for measuring the surface tension of a liquid is the one described in the previous communication as the bubble method. In this, the tube is filled to near its open end, so that a hemispherical meniscus lies at a short distance from that end, as shown in fig. 3.

On setting the table in motion the meniscus extends, and at the critical speed (which is lower than that appropriate to the case represented by fig. 1) the lip of the meniscus curls inwards at the open end of the tube and

Fig. 2.

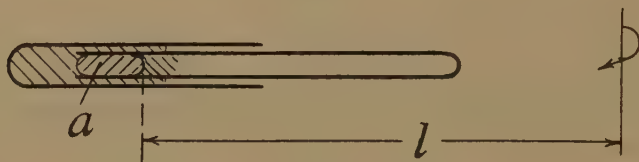
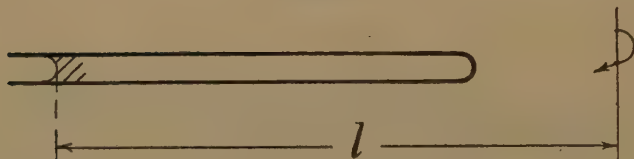


Fig. 3.



encloses an air bubble. With a further slight increase of speed the bubble moves slowly towards the closed end of the tube and comes to rest at a position dependent on the rate of rotation. The formula for this case is

$$T = \rho l 0 d^3 \pi^2 n^2 l. \quad \dots \dots \dots (2)$$

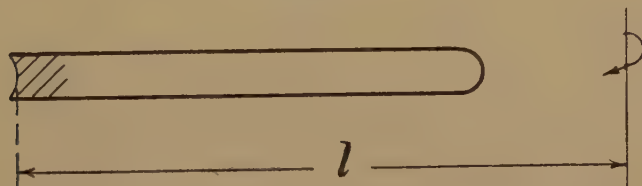
It will be noticed that we have stressed that in each of the variants above described, the initial form of the meniscus must be hemispherical. But not all liquids satisfy this condition, for example, such viscous liquids as paraffin liquid or olive oil, and, consequently, as it is difficult to measure exactly the depth of the meniscus, the procedures illustrated above cannot be used for measuring the surface tensions of these liquids. Moreover, we have found that when measuring interface tensions by the procedure represented in fig. 2, the action involved is very sensitive to the condition of the surface of the bore of the inner tube, and the results obtained can be influenced considerably by the reagents used in cleaning the tube. Our experience has therefore revealed a need for a further variant which on the one hand makes it unnecessary to start with a hemispherical meniscus,

and on the other hand renders the action less dependent on the tube-cleaning technique.

For the measurement of the surface tensions of mobile liquids, such as water, chloroform, toluene, acetone, benzene and the like, the bubble method appears to leave nothing to be desired, and, provided ordinary precautions are taken in cleaning the tube, results can be obtained easily and consistently. But when it is required to deal with viscous liquids, or liquids which do not form a hemispherical meniscus, a different procedure is needed, and for the measurement of interface tensions the problem appears to require the reducing to a minimum of the area of contact between the liquids and the inner tube in the region of the meniscus. To meet the above requirements we have devised a variant depending on a meniscus which is initially flat or very shallow, and is located at the open end of the tube as shown in fig. 4.

For measuring the surface tension of any liquid the procedure is the same as that already described with reference to fig. 1. For measuring the interface tension of two liquids the procedure is similar to that represented

Fig. 4.



at fig. 2, excepting that the inner tube is completely filled with the liquid of higher density, and the other liquid is contained entirely in the outer tube, with a flat or very shallow meniscus between them. In each case, however, the critical speed is higher, with a given tube, as compared with the critical speeds of the methods previously described. If the critical speed with a hemispherical meniscus is  $n_0$ , the critical speed ( $n_1$ ) with a flat or very shallow meniscus is  $2n_0$ . This point is referred to again later in the section dealing with the theory of the method.

In the early stages of our work on the shallow meniscus, we started with a perfectly flat meniscus, but we found that on rotating the table a minute quantity of liquid escaped from the tube, leaving a stable and very shallow curved meniscus. It appears to be immaterial whether the meniscus is initially flat or curved, but if curved it must be very shallow. If too deep, the critical speed is no longer  $2n_0$ , but something less. One very simple method of obtaining a suitably curved meniscus, which is easily acquired after a little practice, consists in filling the tube so as to leave a flat surface on the liquid, and then lightly wiping the tube with a single stroke over a flat sheet of writing paper placed on a table. Another method is to employ a tube having secured to its open end a very thin metal disc in which is formed a central hole of slightly less diameter than the tube bore. After filling, excess liquid is removed by wiping the face

of the disc. But in either case it is essential that the end of the tube or the face of the disc shall be very smooth and flat, and that the edge of the bore or hole shall be sharp. The flat or shallow meniscus method enables the surface tension of viscous liquids, such a paraffin liquid and olive oil, to be very easily ascertained. Moreover, it greatly simplifies the measuring of the interface tensions of either mobile or viscous liquids, as the action is less susceptible than the method of fig. 2, to the surface conditions of the tube bore, and eliminates the need for starting with a hemispherical meniscus.

The formula used with the flat or shallow meniscus method is the same as (1), excepting that  $n$  becomes  $n_1/2$  where  $n_1$  is the observed critical speed.

*Surface Tensions of Viscous Liquids.*

Liquids.	$d$ (cm.).	$l$ (cm.).	$\rho$ .	T.	Critical Speed (r.p.m.).	
					Calculated.	Observed.
Paraffin liquid ..	0.2	10	0.87	30	498	500
Olive oil ..	0.2	10	0.92	33	508	510

*Interface Tensions.*

Liquids.	$d$ (cm.).	$l$ (cm.).	$\rho$ (difference).	T.	Critical Speed (r.p.m.).	
					Calculated.	Observed.
Water-Toluene ..	0.32	10	0.13	36.1	894	894
Water-Benzene ..	—	—	0.12	35	914	914
Water-Paraffin oil	—	—	0.2	48.3	834	840
Water-Paraffin liq.	—	—	0.13	47	1020	1020
Water-Olive oil ..	—	—	0.08	20.6	858	858

Using the figures for surface tension and density given in standard tables for the liquids used, the critical speeds were calculated, and these particulars, together with the observed critical speeds, are given in the preceding tables.

As with other methods some preliminary practice is needed, particularly in connection with the measurement of interface tensions. But when the necessary technique has been acquired, consistent and accurate results are easily obtained. When the results are erratic these are usually attributable to a malcondition of the tube bore surface. As examples of results obtained by the shallow meniscus method, the following may be cited.



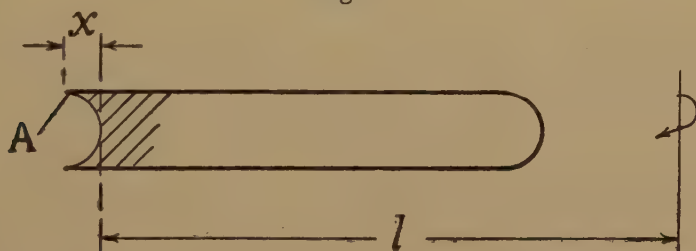
In each case the observed critical speeds in six successive tests are quoted : Interface tension of water-paraffin oil, 840, 840, 840, 840, 850, 840 r.p.m. ; Surface tension of olive oil, 510, 510, 515, 515, 510, 510 r.p.m.

With the liquids above mentioned, when the critical speed is reached, the movement of the liquid or liquids is continuous, and it occurs at a fairly rapid rate. But we have found that when measuring the interface tension of serum and olive oil a quite different effect occurs. In this instance, at the critical speed, the interchange of position of the two liquids occurs as a succession of drops, alternately moving in opposite directions. The same effect occurs between other pairs of liquids having low interface tension and where at least one of the two liquids has high viscosity. Moreover, discharge in the form of drops, instead of a continuous stream, occurs with single liquids having very high viscosity, such as, for example, glycerol.

### THEORY.

The variants of the method forming the subject of our present and previous communications are divisible into two classes (*a*, *b*). In those belonging to class *a* there is no change in the shape of the meniscus until

Fig. 5.



the critical speed is reached. The only example known to us at present in class *b* is the bubble method. In this the meniscus changes its form before the critical speed is attained. The theory is based on the assumption that at the critical speed the pressure per unit area due to surface tension at the region A in the meniscus (fig. 5), is balanced by the pressure due to centrifugal force associated with the length *x* of the meniscus.

The general equation which appears to cover all variants is

$$T \left( \frac{1}{R_1} + \frac{1}{R_2} \right) = \rho x 4\pi^2 n^2 l, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where  $R_1$   $R_2$ =radii of curvature of the meniscus at A in two planes at right angles.

$\rho$ =density of liquid (or difference of densities of two liquids in contact).

$x$ =length of meniscus.

$n$ =critical speed.

$l$ =distance of base of meniscus from axis of rotation.

In the examples shown in figs. 1 and 5, where the meniscus is hemispherical and is situated at the open end of the tube,  $R_1=R_2=r$  (the radius of the tube bore), and  $x=r$ . Formula (3) then becomes

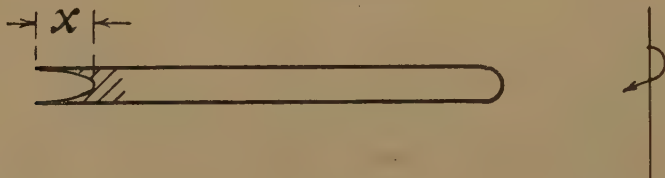
$$T = \frac{1}{2} \rho d^2 \pi^2 n^2 l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

which is the formula given in the previous communication. When the meniscus is not initially hemispherical, but has a shallower form, the critical speed  $n_1$ , at which discharge occurs is higher than  $n_0$  (for a given tube), and depends on  $x$ . If the meniscus has a spherical form in which  $R_1=R_2=R$ , formula (3) becomes

$$T = \rho \frac{Rx}{2} 4\pi^2 n^2 l . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The quantity  $Rx$  tends towards the limiting value  $r^2/2$ , so long as the meniscus closely approximates to a spherical form. In fact this occurs when  $x$  is about  $r/2$ , and in this case the ratio  $n_1/n_0$  is  $\sqrt{2}$ . But when the meniscus is very shallow (and is coincident with the tube end) experiments show that the ratio  $n_1/n_0$  becomes 2, from which it is to be inferred that the meniscus is no longer spherical but has a shape of varying curvature.

Fig. 6.



In the bubble method the meniscus is free to move when the tube is rotated, and just prior to enclosing the air bubble it has an elongated form, as indicated at fig. 6.

It is reasonable to assume that  $1/R_2$  in formula (3) can be neglected, and as  $R_1=r$  the formula becomes

$$T = \rho r x 4 \pi^2 n^2 l. \dots \dots \dots (6)$$

The evaluation of  $x$  was effected by experiments with different liquids (of known  $T$  and  $\rho$ ) in tubes of different diameters, ranging from 0.1 to 0.3 cm. bore. The results gave  $x=20r^2$ . Formula (6) then becomes

$$T = \rho l 0 d^3 \pi^2 n^2 l. \quad (7)$$

as already given in our previous communication.

The experiments yield no information as to the shape of the meniscus at the critical speed. They merely give a measure of  $x$  at that speed. It is, however, a reasonable assumption that the meniscus is then of semi-ellipsoidal form, and from this is deducible the interesting conclusion that the radius of curvature at the base of the meniscus is constant for all liquids and all tubes (within the range of the experiments), the curvature being  $x/r^2=20$ . At first sight it may be thought that formula (7) is

unsound as its dimensions do not balance. Actually this is not the case as the constant 10 is a dimensional quantity and not a mere number.

On comparing formulæ (1) and (2) and denoting the critical speeds as  $n_0$  and  $n_1$  respectively, it will be seen that these speeds are related by the ratio  $n_0^2/n_1^2=20d$ . It follows that if  $n_0=n_1$ , then  $d=0.05$  cm., from which it is to be inferred that in a tube of this diameter the bubble would not form, as the speed required is then the same as the "emptying" speed. We have not been able to confirm this inference owing to the difficulty of observing liquid movements in a tube of such small bore. For a tube of 0.1 cm. bore  $n_1^2/n_1^2=2$ , and for a tube of 0.2 cm. bore  $n_0^2/n_1^2=4$ . Both of these results have been confirmed by experiments.

As mentioned in our previous communication, a disadvantage of the method lies in the fact that the critical speed appears in the formulæ as  $n^2$ , and a small error in the measurement of  $n$  may therefore result in a relatively large error in the calculation of  $T$ . Hitherto we have used an apparatus in which the table is driven by a variable speed electric motor, and the accurate regulation of the speed has not always been easy. The difficulty could probably be overcome by using an A.C. synchronous motor for driving the table at a known and constant speed, and arranging for  $l$  to be varied by movement of the tube radially along the table while the latter is rotating.

Our present experience leads to the conclusion that the flat or shallow meniscus method is the one which is most convenient for measuring interface tensions, and the surface tensions of viscous liquids, or of mobile liquids whose contact angle is not zero. For measuring the surface tension of other mobile liquids the bubble method is the one which gives the most consistent results. All the variants described are especially useful where only small quantities of liquid are available.

---

## VII. *A Non-Quantum Indication of Planck's Radiation Formula.*

By HERBERT DINGLE, Professor of Natural Philosophy,  
Imperial College of Science and Technology \*.

[Received February 8, 1946.]

### *Introductory.*

IN a former paper† the relations between black body radiation and temperature were re-expressed in terms conformable with the fact that we do not observe radiation from a single body, but only what we describe as an exchange of radiation between bodies at different temperatures.

---

\* Communicated by the Author.

† Phil. Mag. xxxv. p. 499 (1944).



The treatment was analogous to that now generally acknowledged to be necessary in the study of the motion of bodies, and demanded there by the fact that we do not observe the motion of a single body, but only the motion of one body with respect to another. Only constant temperatures were considered, corresponding to the restriction to constant velocities in the special theory of mechanical relativity, and the significance of the theory as then developed lay (except from the æsthetic point of view), not so much in the results arrived at, as in the opening up of a new direction of approach to the problem of radiation and the indication of the next steps to take in the development of the subject. Two phenomena—namely, the efficiency of a perfect heat engine and the general equations of the thermo-electric circuit—were indeed mentioned as possibly falling within the scope of this “special” theory of thermal relativity, and it was shown that they satisfied the condition of invariance to transformations of temperature of the instruments employed when these were properly chosen, but I had not the wit at the time to see that by taking account of the fact of observation that black body radiation is associated with a continuous range of periodicities, the theory might afford a simple means of determining how the radiation was distributed among the various frequencies. This has now been done, and it appears that the equivalent of the Planck and not that of the Rayleigh-Jeans formula fulfils the condition of invariance to change of “coordinate systems.”

### *Outline of the Theory.*

It is first desirable very briefly to summarize the essential points of the theory and to reproduce the chief formulæ for convenience of reference. The procedure consists in expressing the phenomenon of radiation in terms corresponding to those employed for expressing the phenomenon of motion. We measure motion (kinematically) by velocity defined as rate of change of space with time. We measure space by a chosen instrument, and time by the space covered by a standard moving body whose velocity is postulated to be uniform (in relativity the moving body is a beam of light; in Newtonian mechanics it is a body free from forces: the time scales thus obtained are identical). The velocity of a body is the space it covers in unit time. Accordingly, we seek to measure radiation by temperature defined as rate of change of some measurable characteristic of the radiating body with time. The characteristic we choose is the value of  $\eta = \beta \log_e p_2/p_1$  when the radiation of the body is absorbed isothermally by one gram-molecule of a perfect gas whose initial and final pressures are  $p_1$  and  $p_2$  respectively;  $\beta$  is an arbitrary constant to determine a convenient value for the unit of  $\eta$ . Time ( $t$ ) we measure by the  $\eta$  received from a standard radiating body, which we choose as one gram-molecule of a perfect gas kept at a slightly higher pressure than, and having the same initial volume as, the gram-molecule of gas which measures the  $\eta$  received from it. This combination of one gram-molecule of a perfect gas radiating continuously to another constitutes

our clock, just as the ideal mechanical clock consists of a beam of light moving continuously along a space-measuring scale. To examine the radiation of a body at constant temperature, then, we bring it up to a separate  $\eta$ -measuring instrument and record the  $d\eta$  received by the instrument in a time  $dt$  measured by the clock, the temperature of the  $\eta$ -measuring instrument and that of the receiving gram-molecule in the clock being, of course, the same, just as the velocity of the scale for measuring space and that of the space-measuring scale forming part of the ideal clock are the same in mechanics. The temperature of the radiating body is then defined as  $\tau = d\eta/dt$ .

If, now, we change to a pair of measuring instruments having a different temperature (say  $\tau_1$ , the temperature of the original instruments being 0), and examine the radiation of the same body with them, we shall have different values of  $d\eta$  and  $dt$  (viz.  $d\eta'$  and  $dt'$ ) and a different value,  $\tau'$ , for the temperature. The known fourth-power law of radiation and the perfect gas equation (which, of course, express experimental facts, in themselves independent of the terms in which we choose to express them) enable us to write down the laws of transformation of  $d\eta$ ,  $dt$  and  $\tau$ , and also the relation of these quantities to the ordinary measures of temperature, time, etc. The relevant equations are as follows :

$$\left. \begin{aligned} d\eta &= -\frac{\beta a}{R\theta_0} (\theta^4 - \theta_0^4) d\bar{t}, \\ dt &= -\frac{\beta a \theta_0^3}{R\zeta} d\bar{t}, \\ \tau &= \zeta \left( \frac{\theta^4}{\theta_0^4} - 1 \right). \end{aligned} \right\} \dots \dots \dots (18')$$

$$\left. \begin{aligned} d\eta' &= \frac{d\eta - \tau_1 dt}{\psi_1^{\frac{1}{2}}}, \\ dt' &= \psi_1^{\frac{1}{2}} dt, \\ \tau' &= \frac{\tau - \tau_1}{\psi_2}. \end{aligned} \right\} \dots \dots \dots (21')$$

Here, as elsewhere, equations which appear in the former paper have been given the same number with a dash, to facilitate reference.  $R$  is the gas constant ;  $a\theta^4$  the energy radiated per second by the body under examination at Kelvin temperature  $\theta$ , so that  $a$  is Stefan's constant multiplied by the area of the body ;  $\theta_0$  is the Kelvin temperature of the instruments ;  $d\bar{t}$  is the duration of the radiation according to the ordinary mechanical clock of current theory ; and  $\zeta$  is a constant whose value depends on the unit of thermal time chosen.  $\zeta$  corresponds on the  $\tau$ -scale of temperature to the absolute zero of the Kelvin scale ; its analogue in mechanical relativity is the velocity of light.

There is one formal difference between equations (21') and (21) in the former paper, namely, that  $\psi_1$  has been written for  $1 + \tau_1/\zeta$ . Since this





any such theory. It is a feature, in fact, of the present treatment that it is independent of such theories, whereas on the ordinary view of radiation we cannot escape the inconsistency that the matter-like ether shown to be impossible in mechanics is found to be indispensable for radiation. Here we merely assert that the characteristic of the body under examination, which we usually call that of radiating energy, is a periodic phenomenon, and that means, in terms of the theory, that when we change our "co-ordinate system" (*i. e.* the temperature of our measuring instruments) we must transform the frequencies concerned according to the transformation of thermal time given in equations (21'). The frequencies so obtained, of course, will be quite different from those normally assigned to the radiation, but it is an essential and generally acknowledged (though less generally understood \*) element of relativity theory that our assignments of time intervals have the character of co-ordinates which we can change freely according to a chosen principle. If we denote frequency on the thermal scale by  $\nu$ , then obviously the transformation is the inverse of that for time, viz.

$$\nu' = \psi_1^{-\frac{1}{2}} \nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Let us, then, expose our radiating body, at constant temperature  $\tau$ , to the  $\eta$ -measuring instrument at some arbitrary zero temperature for a definite duration measured by the thermal clock as  $dt$ , and by the ordinary mechanical clock as  $\tilde{dt}$ , and let us insert *ad lib.* appropriate filters between the body and the instrument so as to restrict the observations to any arbitrary frequencies we please. We can then measure  $d\eta$  and  $dt$  and so evaluate  $d\sigma$  for each frequency. The value of  $d\sigma$  so obtained will be a function of  $\tau$  (or  $\psi$ ), of the duration  $dt$ , and of the frequency  $\nu$  with which we are concerned, so that we can write for the unfiltered radiation

$$d\sigma = \int_0^\infty F(\psi, \nu, t) d\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Obviously  $d\sigma$  will be proportional to  $dt$ , so we can re-write this as

$$d\sigma = d' \int_0^\infty f(\psi, \nu) d\nu. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$f(\psi, \nu)$  is clearly the amount of  $d\sigma$  per unit time per unit range of  $\nu$ . If the instruments had been at some other temperature,  $\tau_1$ , instead of zero we should have had similarly

$$d\sigma = dt' \int_0^\infty f(\psi', \nu') d\nu', \quad (6)$$

and the invariance of  $d\sigma$  requires that

$$dt \int_0^\infty f(\psi, \nu) d\nu = dt' \int_0^\infty f(\psi', \nu') d\nu'. \quad (7)$$

\* See, for instance, 'Nature,' cxliv. p. 888 (1939); cxlvi. p. 391 (1940), and associated correspondence; also Amer. Journ. Phys. x. p. 1 (1942); x. p. 203 (1942); xi. pp. 219, 228 (1943).

Hence, using the transformation equations (21'), (2) and (3), we have

$$dt \int_0^\infty f(\psi, \nu) d\nu = \psi_1^{-\frac{1}{2}} dt \int_0^\infty f(\psi \psi_1^{-1}, \psi_1^{-\frac{1}{2}} \nu) \psi_1^{-\frac{1}{2}} d\nu;$$

i. e.  $\int_0^\infty f(\psi, \nu) d\nu = \int_0^\infty f(\psi \psi_1^{-1}, \psi_1^{-\frac{1}{2}} \nu) d\nu. \quad \dots \dots \dots (8)$

The function we require must satisfy this relation, must be independent of  $\psi_1$ , and must make the integrals equal to  $\frac{d\sigma}{dt}$ , i. e.  $\psi^{\frac{1}{2}}$ .

Consider the function

$$f(\psi, \nu) = \frac{A \nu^3 \psi^{-\frac{1}{2}}}{e^{B \nu \psi^{-\frac{1}{2}}} - 1}, \quad \dots \dots \dots (9)$$

where A and B are arbitrary constants. On substitution, the l.h.s. of (8) becomes

$$A \psi^{-\frac{1}{2}} \int_0^\infty \frac{\nu^3 d\nu}{e^{B \nu \psi^{-\frac{1}{2}}} - 1},$$

the value of which is well known to be

$$A \psi^{-\frac{1}{2}} \left( \frac{1}{B^4 \psi^{-1}} \times \frac{\pi^4}{15} \right) = \frac{\pi^4}{15} \frac{A}{B^4} \psi^{\frac{1}{2}}. \quad \dots \dots \dots (10)$$

The factor  $\pi^4/15$  is six times the series  $1/1^4 + 1/2^4 + 1/3^4 \dots ad. inf.$  The r.h.s. of (8) is

$$A \psi_1^{-2} \psi^{-\frac{1}{2}} \int_0^\infty \frac{\nu^3 d\nu}{e^{B \psi_1^{-\frac{1}{2}} \psi^{-\frac{1}{2}} \nu} - 1},$$

of which the integral is similarly

$$A \psi_1^{-2} \psi^{-\frac{1}{2}} \left( \frac{1}{B^4 \psi_1^{-2} \psi^{-1}} \times \frac{\pi^4}{15} \right) = \frac{\pi^4}{15} \frac{A}{B^4} \psi^{\frac{1}{2}}. \quad \dots \dots \dots (11)$$

The function (9), therefore, satisfies the conditions imposed if

$$A = \frac{15 B^4}{\pi^4} \dots \dots \dots (12)$$

It then becomes

$$f(\psi, \nu) = \frac{\frac{15 B^4}{\pi^4} \nu^3 \psi^{-\frac{1}{2}}}{e^{B \nu \psi^{-\frac{1}{2}}} - 1} \dots \dots \dots (13)$$

### Correspondence with the Planck Formula.

Planck's well-known formula gives the distribution of energy among the frequencies according to the ordinary mechanical scale of time measurement: (13) gives the distribution of  $d\sigma$  among the frequencies according to the thermal scale of time measurement. It is easy to show, however, that the two formulæ are precisely equivalent. The expression (13) is the value of  $d\sigma$  per unit thermal time per unit range of thermal frequency  $\nu$ . It is also the value of  $d\sigma$  per unit mechanical time per unit

range of mechanical frequency  $\bar{\nu}$ , for in order to convert it into these terms we must multiply the expression by  $\frac{d\bar{t}}{dt} \cdot \frac{\bar{\nu}}{\nu}$ , which clearly leaves it unchanged. Substituting from (18'), so as to transform it into ordinary terms, we have therefore

$$\left. \begin{array}{l} d\sigma \text{ per unit mechanical time per unit} \\ \text{range of mechanical frequency} \end{array} \right\} = \frac{\frac{15B^4}{\pi^4\theta_0^3} \left(\frac{R\zeta}{\beta a}\right)^3 \bar{\nu}^{3\theta-1}}{e^{-\frac{BR\zeta}{\beta a\theta_0^3} \bar{\nu}^\theta} - 1}. \quad (14)$$

Now, from (24') the *energy* radiated per unit mechanical time is obtained by multiplying  $d\sigma$  by  $-R\zeta\theta/\beta$ , and if we consider 1 sq. cm. of the radiating surface, so that  $a$  becomes Stefan's constant, we obtain

$$\left. \begin{array}{l} \text{Energy radiated per unit area per} \\ \text{unit mechanical time per unit range} \\ \text{of mechanical frequency} \end{array} \right\} = \frac{15 \left(\frac{BR\zeta}{\pi\beta\theta_0^2}\right)^4 \frac{1}{a^3} \bar{\nu}^3}{e^{-\frac{BR\zeta}{\beta a\theta_0^3} \bar{\nu}^\theta} - 1}. \quad (15)$$

The Planck formula for this is

$$\frac{\frac{2\pi h}{c^2} \bar{\nu}^3}{e^{\frac{h\bar{\nu}}{k\theta}} - 1}, \quad \dots \dots \dots (16)$$

so that the two expressions become identical if we write

$$\left. \begin{array}{l} 15 \left(\frac{B\zeta}{\beta\theta_0^2}\right)^4 = \frac{2\pi^5 h a^3}{c^2 R^4}, \\ -\frac{B\zeta}{\beta\theta_0^2} = \frac{ah}{kR} \end{array} \right\} \dots \dots \dots (17)$$

Eliminating  $\frac{B\zeta}{\beta\theta_0^2}$  we obtain

$$a = \frac{2\pi^5 k^4}{15 h^3 c^2} \dots \dots \dots (18)$$

This is, of course, a well-known result, and its derivation here serves merely as a check on the calculation.

It is interesting to note that the Rayleigh-Jeans formula, viz.

$$E_\nu = \frac{2\pi k \bar{\nu}^2 \theta}{c^2}, \quad \dots \dots \dots (19)$$

which corresponds in our terms to  $d\sigma = G\nu^2$ , where  $G$  does not involve  $\psi$  or  $\nu$ , does not satisfy the conditions of invariance, and is therefore inadmissible.

It will be noticed that the right-hand side of equations (17) contains only quantities significant in current theory, and the left-hand side only quantities significant in the present theory. Of the latter,  $\beta$  and  $\zeta$  merely indicate our freedom of choice of units of  $\eta$  and thermal time





*Significance of the Result.*

The significance of the result arrived at may be summarized as follows. First, a new condition, not previously known, has been imposed on the function representing the distribution of energy in the spectrum of a black body. Since the Rayleigh-Jeans law does not satisfy this condition it is seen to be inadmissible quite apart from comparison with observation, and as it is a necessary consequence of the classical atomic theory the new condition requires the rejection of that theory. This is an example of the necessary subservience of microscopic to macroscopic laws which seems to me incompatible with the view commonly held that the latter are simply the approximate forms taken by the former when the number of units involved is very large.

Secondly, the conclusion, which Jeans \* attributes to Poincaré, that "the mere fact that the total radiation at a finite temperature is finite requires that the ultimate motion should be in some way discontinuous," is no longer valid. The finiteness of the total radiation is a condition satisfied by the Planck formula independently of any hypothesis of the "ultimate motion." If one wishes to advance from equation (8), with its associated conditions, to the solution which agrees with observation, he can do so either by inventing hypotheses until he finds a successful one or by testing functions until he does so. The former process is at least as arbitrary as the latter, and even less definitive, because there must always remain the possibility that some other hypothesis would lead to the same formula. The Planck formula therefore by no means demands any conception of discontinuity; on the contrary, it is expressed in the form (13) in a way which involves only continuous functions of observed quantities.

Thirdly, the fact that the new condition is satisfied by the Planck and not by the Rayleigh-Jeans law is evidence of the soundness and usefulness of the principles on which it is based. These are (1) the invariance of  $d\sigma$ ; (2) the postulate that if light is regarded as a periodic phenomenon, the period is subject to the transformation law for thermal time. The invariance of  $d\sigma$  involves nothing new except the form in which it is expressed; as was shown in the former paper †, it follows by mathematical necessity from the ordinary fourth-power law of black body radiation and the properties of perfect gases. So regarded, the quantity  $dt$ , it is true, is simply a curious function of time and temperature measurements, but the mathematical relation is unquestionable. The point of view of the thermal relativity theory is necessary to reveal  $dt$  as a simple measure of time, alternative to and co-valid with that of current mechanical theory. The fact, then, that the postulate that the frequency of light is subject to the transformation law for  $dt$  leads to the Planck and away from the Rayleigh-Jeans law gives strong support

---

\* 'Radiation and the Quantum Theory' (Second Edition), p. 29.

† *Ibid.*, pp. 511-14, in the section headed "Mathematical Aspect of the Theory."

both to the fundamental principles of thermal relativity and to the view that it may prove a useful instrument of investigation. Confirmation is also given of the general view of time measurement which I have previously put forward\* on more fundamental grounds. It does not seem out of place, therefore, to consider in a preliminary way what possibilities exist of extending its field of application.

### *Possibility of a New Thermodynamics.*

The most immediate development would seem likely to be the combination of relativistic mechanics with the new treatment of radiation, so as to form a system of thermodynamics on different lines from the existing one†. We are already familiar with two alternative systems of thermodynamics—the macroscopic, so-called “classical,” thermodynamics, in which union is made through the common concept of energy, and statistical mechanics, in which the concept of atoms is the connecting link. If we attempt to combine relativistic mechanics (which, by its inclusion of gravitation, covers an enormous field so far untouched by microscopic theory) with the relativistic theory of radiation, the natural connecting link would seem to be time, which is common to both schemes. There is an immediate difficulty, however, in the fact that the two systems of time measurement are non-conformable, so to speak. The relation between them, given by the second equation of (18'), involves  $\theta_0$ , the arbitrary zero of temperature, and this means that in order to relate the two time systems we must first choose a value for  $\theta_0$  and also (though, of course, the necessity does not appear in our equations since we have had no need to introduce it) an arbitrary standard of rest to give precision to the unit of mechanical time. We can, it is true, choose some natural phenomenon, such as the Earth, to define a standard of zero motion, and some other natural phenomenon, such as ice melting at normal atmospheric pressure, to define a standard of zero temperature, and the two time scales could then be definitely related with one another by making the units at the standard conditions identical. Such choices, however, would be purely arbitrary, and could not possibly yield a thermodynamics satisfying the primary relativistic requirement of covariance. If, on the other hand, we proceed to the absolutes of the theories—the velocity of light and the “absolute zero” of temperature—both scales vanish.

If time fails us, the First Law of thermodynamics might be the means of forming the connection. Mechanical energy would then need a thermal analogue which might perhaps be constructed from the following

\* ‘Nature,’ cxliv. p. 888 (1939); cxlvi. p. 391 (1940).

† The possibility of such a system of thermodynamics arising from a thermal measure of time was forecast much earlier (see the author’s “Through Science to Philosophy,” O.U.P., 1937, pp. 298 *et seq.*). At that time, of course, a thermal time scale had not been constructed but merely envisaged on general rational grounds as a legitimate possibility.



considerations. In special relativity mechanics the idea of mass ( $M$ ) is introduced as a quantity which satisfies the laws of conservation of itself and of momentum. Such a quantity, as is well known, is found to be a function of velocity and to have the value  $M_0c^2$ , in energy measure, at the arbitrary position of rest. A similar quantity  $H$  (with obvious resemblance to heat capacity) might be defined and expressed as a function of  $\tau$  by considering the exchange of heat between bodies as a "thermal collision." The condition that this quantity and its product with a function of  $\tau$  must be conserved in such a collision would express it as a function of  $\tau$ , and at the arbitrary zero of temperature we might expect it also to appear in "energy" measure as a function of  $H_0$  and  $\zeta$  (say  $H_0\zeta^2$  for purposes of illustration). We might then identify  $M_0c^2$  and  $H_0\zeta^2$  as different expressions of the same characteristic—the "energy of constitution," as it is often called—of the body in question.

If such a program were carried out it would automatically yield us an expression for the variation of  $H$  with temperature which would express the same facts of experiment as our present formulæ for the variation of specific heat with temperature, which fall within a well-known field of application of quantum theory. The inexactness and complication of the formulæ which have been derived, however, indicate clearly enough that specific heat is a very unsuitable conception for theoretical treatment (although, of course, probably indispensable in engineering), and the difficulty in the relativistic treatment would lie not so much in defining the quantity  $H$  and deducing its relations with temperature as in connecting it with the specific heat of ordinary theory. The  $\eta$  which we measure is (to speak in ordinary terms) determined by the surface of the body, and our quantity  $H$  would therefore be related not directly to specific heat but to some function of specific heat, thermal conductivity, ratio of surface to volume, etc., and would, moreover, in its definition presuppose that all mechanical accompaniments of temperature changes were allowed for, just as in the treatment of mechanical collisions we tacitly assume that no energy is used in heating the colliding bodies. The latter assumption is realized nearly enough, but the former is far from being realized. Furthermore, it is doubtful whether such a treatment of thermal collisions can properly be conducted in terms of a "special" theory. In mechanics it is often possible to neglect gravitational fields because we can easily maintain the velocities of bodies approximately constant. In heat the strong tendency of bodies to come to the same temperature is equivalent to an enormous thermal "gravitational field," and it might well be that a generalization of the theory is necessary before it becomes profitable to consider the exchange of heat between bodies at different temperatures.

In principle, however, it seems reasonable to suppose that the difficulties of establishing a general relativistic science of thermodynamics are not insuperable. If they are overcome the applications to cosmology will be obvious and important; the scope of the Second Law of thermodynamics might be more fully understood, and much of the present

difference of opinion concerning its implications removed. This law might, indeed, form the means of creating the link we are seeking, since our invariant  $d\sigma$  corresponds to entropy, but inasmuch as the Second Law is not at present metrical (it merely says that entropy cannot decrease, but gives no measure of its rate of increase), the First Law seems more promising for this purpose.

### VIII. *A Relativistic Determination of the Relation between Electrical Resistance and Temperature.*

By HERBERT DINGLE, Professor of Natural Philosophy,  
Imperial College of Science and Technology \*.

[Received February 18, 1946.]

THE relativistic theory of temperature radiation which I recently put forward †, and which is summarized in the preceding paper, is primarily significant as a method of correlating thermal phenomena which is alternative to, and epistemologically more satisfactory than, that which implies the existence of unobservable absolute radiation in a constant temperature system. As already shown ‡, it imposes certain restrictions on the frequency distribution in the radiation passing between bodies at different temperatures, which make inadmissible the Rayleigh-Jeans law, but allow the law of Planck which is known to agree with observation. The purpose of this paper is to show that the theory may be used as an instrument for deriving physical relations not immediately connected with temperature radiation, just as the second law of thermodynamics, in itself a statement concerning the relations between heat and work, may be used to derive relations in other fields such as those of chemistry and electricity. The result derived here is the dependence of the electrical resistance of a pure substance on temperature; the theory not only requires that there is such a dependence, but also indicates what its character must be.

The general principle is as follows:—Consider a conducting circuit in which a constant e.m.f. is introduced and maintained. A current will flow, and the temperature of the circuit will rise until, in the absence of any other influences, the rate of generation of Joulean heat is equal to the rate of radiation from the circuit to the surroundings. Thereafter the circuit will be in thermal equilibrium at a Kelvin temperature  $\theta$ , the temperature of the surroundings being  $\theta_s$ , and we shall have the relation

$$\overline{Ei} = a_1(\theta^4 - \theta_s^4), \quad . . . . . (1)$$

\* Communicated by the Author.

† Phil. Mag. xxxv. p. 499 (1944).

‡ Phil. Mag. xxxvii. p. 47 (1946).

where  $\bar{E}$  and  $\bar{i}$  are respectively the e.m.f. and the current, and  $a_1$  is a constant for the circuit. Now according to the relativistic theory this relation should hold good irrespective of the temperature of the instruments used for measuring the radiation, and if these are of the type prescribed in the theory we can write down the way in which the r.h.s. of (1) will be transformed as a result of changing from instruments at one temperature to those at another. In order to maintain the relation, therefore, the l.h.s. must transform in the same manner, and we shall find that for this to be so the quantity  $\bar{E}\bar{i}$  must be a particular function of the temperature of the circuit. If, then, we choose a method of producing the e.m.f. which we know to be independent of temperature, the dependence on temperature must characterize the current only, and we can then directly deduce the dependence on temperature of the resistance of the circuit.

In considering the equilibrium of the circuit we substitute for our ordinary clock and thermometers the thermal clock and the instruments which measure temperature on the  $\tau$ -scale, but make no change in any other fundamental measurements (*e.g.* those of mass and space in the traditional dimensional system). We must therefore re-express equation (1) in terms of the measurements obtained when these changes are made. Both sides of the equation will be affected, the r.h.s. on account of the change of thermometer and the l.h.s. on account of the change of clock.

There are two distinct restrictions on our freedom in choosing an actual situation to which the above considerations can be applied. In the first place, our result will be of no value unless the circuit is perfectly homogeneous, for we can draw no significant conclusions about the variation of resistance from a variation of current if it remains uncertain which part of the circuit has suffered the change. This precludes the placing in the circuit of any measuring instrument or source of e.m.f. This ban is necessary also because Peltier and other effects would make impossible a definite value of  $\theta$  applicable to the whole circuit unless the circuit consisted throughout of a single substance. The results we obtain will therefore not be applicable to electrolytes, cells, etc., but chiefly to pure metallic conductors.

Secondly, we must be able to describe the processes of measuring  $\bar{E}$  and  $\bar{i}$  in sufficient detail to be able to say precisely how they involve the measurement of time; otherwise we shall not be able to re-express  $\bar{E}$  and  $\bar{i}$  in the new terms. It may be thought that the dimensions of  $\bar{E}$  and  $\bar{i}$  would give us the necessary information, but this is not so because, in accordance with the principles discussed in previous papers \*, dimensions describe the measuring processes employed, and since there are alternative processes for measuring what is conventionally called the same quantity, there are alternative dimensions. The principle of dimensional homogeneity is, of course, satisfied whatever method is

\* Phil. Mag. xxxiii. p. 321 (1942); xxxiii. p. 692 (1942); xxxiv. p. 588 (1943).



chosen, because the dimensional constants (*e. g.*,  $a_1$  in equation (1)) which inevitably occur in the expression of experimentally determined relations are given whatever dimensions are necessary to satisfy it. Here we are not concerned merely with balancing the dimensions on the two sides of the equation: we need to know the separate dimensions of  $\bar{E}$  and  $\bar{i}$ , and before we can know them we must state exactly what measurements we are going to make to determine the quantities. This seems to me unquestionable, but it is perhaps satisfactory, since there are apparently some who still adhere to the old idea of absolute dimensions, that the occurrence of the "unknown" dimensions of  $\mu$  or  $\kappa$  rules out any possibility of our employing the traditional symbols in this case. We shall return to this point in the last paragraph.

The following procedure, I think, meets the conditions imposed by these two sets of considerations; its practical convenience (or inconvenience) of course is quite immaterial so long as, if difficulties of manipulation are overcome, it tells us unambiguously how the measurement of time is involved and enables us unambiguously to deduce the variation of resistance of a definite substance with temperature.

Suppose a number of exactly similar parallel horizontal magnets are arranged with their N-poles in a vertical plane lying in the Earth's magnetic meridian, and suppose fresh magnets are added to the system at a constant rate so that the field at a horizontal distance  $d$  from the plane, and in a direction perpendicular to it, also grows at a constant rate. At a distance  $2d$  let there be a similar parallel plane containing an exactly similar distribution of N-poles of magnets pointing in the opposite direction, and let this distribution also be added to at the same rate as the other, so that whenever a pole is placed in the first plane an equal pole is also placed at the opposite point in the second. In these circumstances the field at the position of the vertical plane half-way between the two distributions will no longer grow, but will be constant and will consist only of the Earth's field, while the field perpendicular to any other vertical plane between the two systems will grow at a constant rate proportional to that of the systems of poles. Now place the homogeneous coil to which equation (1) is to be applied—a continuous ring of metal wire—in such another vertical plane. The flux,  $B$ , through it will then grow at a constant rate, so that a constant e.m.f. will be induced in it numerically equal to  $dB/dt$ . (The notation of the former papers is used throughout.) Now let a magnetic needle be placed on the axis of the coil and in the plane half-way between the two distributions of poles, where, apart from that due to the current induced in the coil, the only magnetic field is the constant field of the Earth. The needle will then take up a position inclined at  $\phi$  to the meridian such that the induced current will be measured by  $k \tan \phi$ , where  $k$  is a constant.

With this arrangement we have a homogeneous circuit which we can maintain at a uniform temperature, and we can see without any uncertainty how the measurements are affected by a change of instruments. There are few, if any, alternative arrangements which have these

characteristics. The product  $\bar{E}\bar{i}$  is given by  $\frac{dB}{d\bar{t}} \cdot k \tan \phi$ , and if, in determining its value, we decide to change our clock, the only one of our measurements to be changed will be  $d\bar{t}$  in the denominator. Hence, if we denote by  $E$  and  $i$  the values of  $\bar{E}$  and  $\bar{i}$  when the change is made from the ordinary instruments to the  $\eta$ -measuring instrument and thermal clock at temperature  $\theta_0$ , we shall have

$$E \equiv \frac{dB}{dt} = \bar{E} \frac{d\bar{t}}{dt} = \bar{E} \frac{R\zeta}{\beta a \theta_0^3} \text{ numerically } \left. \vphantom{\frac{dB}{dt}} \right\} \quad (2)$$

and

$$i = \bar{i}$$

equation (1) therefore becomes

$$\left. \begin{aligned} E i \frac{\beta a \theta_0^3}{R \zeta} &= a_1 \theta_0^4 (\psi - \psi_s), \\ E i \frac{\beta a}{R \zeta} &= a_1 \theta_0 (\psi - \psi_s). \end{aligned} \right\} \quad (3)$$

*i. e.*

If now we use instruments at Kelvin temperature  $\theta_1 (= \tau_1$  on the previous scale) and denote by  $\psi'$ ,  $\psi'_s$  the temperatures of the coil and its surroundings as measured with these instruments, and by  $E'$ ,  $i'$ , the new e.m.f. and current, we shall have similarly

$$E' i' \frac{\beta a}{R \zeta} = a_1 \theta_1 (\psi' - \psi'_s). \quad (4)$$

Now (3) and (4) have the same form only if  $E$  or  $i$  varies with temperature so as to eliminate  $\theta_0$  and  $\theta_1$ . Experience shows that  $E$  is independent of temperature, so let us put  $i = I \theta_0$ . Our condition will then be met if  $I$  is invariant, for  $I \theta_0 = I \theta \psi^{-\frac{1}{2}}$ , and this transforms into  $I \theta \psi'^{-\frac{1}{2}}$ , *i. e.*  $I \theta_1$ , so that  $i' = I \theta_1$ .

The dependence of  $i$  on temperature at once follows. We have  $i'/i = \theta_1/\theta_0 = \theta \psi'^{-\frac{1}{2}}/\theta \psi^{-\frac{1}{2}}$ , *i. e.*,

$$i \psi^{\frac{1}{2}} = i' \psi'^{\frac{1}{2}} = i_0, \text{ say,} \quad (5)$$

whence, generally,

$$i = i_0 \psi^{-\frac{1}{2}}. \quad (6)$$

This means that in order that equation (1) shall be true whatever may be the temperature of our measuring instruments, the current must vary with the temperature of the coil according to (6). This is possible only if the resistance,  $\rho$ , of the coil varies with temperature according to the inverse relation

$$\rho = \rho_0 \psi^{\frac{1}{2}}, \quad (7)$$

*i. e.*, in terms of Kelvin temperatures,

$$\rho = \rho_0 \frac{\theta}{\theta_0}. \quad (8)$$





as a means of deriving physical relations, the present method is potentially able to transcend the existing thermodynamical laws because although, like them, it is as yet limited to equilibrium conditions, it contains the possibility of generalization which should make it applicable to changing conditions also. It might then be possible to give as complete, or even more complete, a description of observed phenomena in purely macroscopic terms as we can now with the aid of unobservable microscopic entities. From the epistemological point of view such a result is much to be desired.

In view of the present lack of agreement concerning the theory of dimensions and the character of physical relationships, it is probably not superfluous to anticipate and try to remove an objection which might be raised to the foregoing argument on the ground that the implicit occurrence of time in the conceptions of e.m.f. and current has been ignored. It might be argued, for example, that when we change our unit of time we should change the values assigned to the Earth's field and to the poles of the magnets used, because the definitions of those quantities involve force, which is defined in terms of time among other things. To place the matter beyond question, therefore, we will discard at first all the accepted electric and magnetic measurements and describe the experimental arrangement contemplated, and the relation which it yields, in terms of the actual operations which a person ignorant of those measurements could carry out. Let him set up the apparatus mentioned, using magnets of equal but otherwise unspecified pole strength (he can do this, of course, without requiring any scale of measurement of pole strength), which he brings up at a constant rate as indicated by his thermal clock at Kelvin temperature  $\theta_0(\tau=0)$ . He observes this rate,  $\frac{dn}{dt}$ . He finds that when the temperature of the coil has reached a constant value,  $\tau$  (or  $\theta$ ), the needle is stationary at an angle  $\phi$  with the plane of the coil. He then finds that

$$\theta_0^3 \frac{dn}{dt} \cdot A \tan \phi = \theta_0^4 (\psi - \psi_s), \quad . . . . . (1')$$

where  $A$  is a constant, and that this relation remains true, with the same value of  $A$ , when  $\frac{dn}{dt}$  (and consequently  $\theta$  and  $\phi$ ) is arbitrarily changed, so long as the same measuring instruments are used. He has thus discovered a physical relation, true within one thermal co-ordinate system. If, now, he changes his instruments for measuring temperature and time, without, of course, changing the system set up in any way, then clearly  $\frac{dn}{dt}$  will be different because of the change in  $dt$  for the same value of  $dn$ ;  $\tan \phi$  will be unaffected;  $\theta_0$  will become  $\theta_1$ ; and  $\psi, \psi_s$  will be  $\psi'$  and  $\psi'_s$ . He then finds that a relation of the same form holds good, but  $A$  has become  $A'$ ; *i. e.*, he has a relation

$$\theta_1^3 \frac{dn}{dt'} \cdot A' \tan \phi = \theta_1^4 (\psi' - \psi'_s).$$

This will agree with (1') if

$$\frac{A}{\theta_0} = \frac{A'}{\theta_1} = B, \text{ say,}$$

i. e., if instead of A he writes  $B\theta_0$ . He can then say that his relation (1'), written in the form

$$\frac{dn}{dt} \cdot B \tan \phi = (\psi - \psi_s),$$

is true whatever the temperature of his clock and  $\eta$ -measuring instrument, and therefore this is the invariant form of the relation he has discovered. B is independent of the temperature of the instruments, but A depends thereon; or, in other words, if A represents a physical quantity it must be a function of temperature.

But now, to introduce ordinary electrical terms for the first time, it is clear that A, or  $A \tan \phi$ , is proportional to the *current* in the coil, for it is proportional to the result of a standard method of measuring current, namely, by a tangent galvanometer. Hence the fact that A varies with temperature means that the current must vary with temperature, and this can be so only if the *resistance* varies in the inverse way, for we know that E is independent of temperature. Hence we have obtained our result without using any quantities to which unique dimensions can be assumed to be attached, and deduced its applicability to such quantities by showing that the measurements involved are, in fact, measurements of them; i. e., that our result would not hold good unless the quantities varied in the manner required.

## IX. On the Dimensions of Physical Magnitudes (Sixth Paper).

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.\*

[Received May 6, 1946.]

MR. DALZELL gives a statement † of the principles of measurement to which he is, of course, perfectly entitled to adhere, but he should not have placed it in relation to mine, as he has done, without stating that he uses the word "magnitude" in a quite different sense. Mr. Dalzell pictures a "physical quantity" as a sort of objectively existing amount of stuff from which one can take arbitrarily any portion and call it a unit. The "magnitude" of the unit he conceives as the amount of stuff in this portion, and the amount of stuff in any other portion (the "magnitude" of that portion of the physical quantity) stands to the magnitude of the unit in a certain numerical ratio which the process of measurement

\* Communicated by the Author.

† Phil. Mag. xxxvi. p. 485 (1945).

discovers. Measurement thus never determines any single magnitude but only a ratio of magnitudes of two samples of the quantity, of which one (the unit) is arbitrary. This seems to me the only possible interpretation of such a statement as "Given a collection of uniformly measurable physical quantities mutually related by physical laws, a particular magnitude of each such quantity, called the unit magnitude, may be selected and given a name. In that case the ratio of a magnitude of one of these physical quantities to the corresponding unit magnitude is called the numerical measure of that magnitude relative to the unit."

The word "magnitude" as I use it, however (see, for instance, definition (3) on p. 322 of my first paper\* on the subject), represents the result obtained after performing a specified operation. It does not rest on the supposed existence, independently of measurement, of a stuff-like physical quantity; the operation comes first (including usually both manual and mental parts) and its result is given the name "magnitude." Throughout my papers I have repeatedly emphasized the inseparability of magnitude and operation of measurement and have even pointed out (pp. 323-4 of the same paper) that the notion of a physical quantity could be dispensed with altogether without affecting anything essential. With my definition a particular magnitude means something precise and something uniquely related to the particular example to which it refers. Thus, to say that the magnitude of the mass of a certain body is 5.64 kg. is to say that when the prescribed operation of measuring mass is performed on that body the result is 5.64 kg. To Mr. Dalzell it is to say that the amount of stuff in that body is 5.64 times as much as the amount of stuff in a certain arbitrary standard body. My objection to this is that it implies that there is an amount of stuff in each body—sample and standard—which is eternally unknowable, and so one has to drag about a collection of metaphysical inaccessible conceptions which demands a practically unattainable perfection of vigilance if one is to avoid being misled by false problems. The position is similar to that existing before the principle of relativity was enunciated. Two different bodies were each supposed to have an unknown absolute velocity: if they were relatively at rest these unknown velocities were the same, but if they were relatively in motion the velocities were different and only the difference was knowable. We all know what a clarification occurred when the unknowable absolute velocities were discarded. I suggest that a similar clarification occurs when the unknowable absolute "magnitudes" of Mr. Dalzell are discarded.

It is indeed surprising how this very important general question of measurement has fallen into a backwater, past which the stream of relativity moves without disturbing traditional and now outmoded notions. Mr. Dalzell writes of Maxwell, for instance, as though his dicta were invulnerable to later criticism. No one who knows anything of physics can easily set bounds to his respect for Maxwell's genius, but to bring forward his theory of measurement as though it now had any authority is

---

\* *Phil. Mag.* xxxiii. p. 321 (1942).



not respect but superstition. The essence of Einstein's great discovery was that we had no operation for measuring the length of a moving body. The supposed operation required one to read the two ends at the same time, and we could not determine uniquely what two moments were at the same time. Hence the length of a moving body was undefined, and we had to make a definition. Maxwell held a view of measurement according to which the moving body had a unique intrinsic length, whether we could measure it or not and whether we could say how to measure it or not. This is quite incompatible with the principle of relativity.

There is one other point on which Mr. Dalzell misrepresents me. For some unstated reason he assumes that when I posed the problem: "Here is a charge; state the operation for measuring it," I was implying that there was no other charge in the world. I was, in fact, taking the world as it is, with no reservation at all about the number of charges it might contain. The problem is to state the operation for measuring charge, and then you know what you mean when you say that a certain charge has a magnitude of 20 units. If you care to change the specification by adding the clause: "Multiply the result by 10," then you change the unit and the magnitude becomes 200. Nothing but convenience need decide whether this should be done or not. Mr. Dalzell seems to think that my prescription would be satisfied by giving the charge a name, such as Archibald. Perhaps it would, though I did say (definition (2) of the paper already cited) that measurement could always in principle involve a numerical process; but life is short, and I did not think it necessary to emphasize that operations of measurement should be such as to yield numbers, any more than to say that they should not normally have to be performed by people standing on their heads. The result may be any number—10,  $19\frac{1}{2}$ ,  $-6\cdot3$ , ...—according to your choice of unit, but Archibald, certainly not!

### *X. Boundary Value Problems of a Heavy Circular Disk held in a Vertical Plane.*

By BIBHUTIBHUSAN SEN \*.

[Received August 28, 1944.]

#### *Introduction.*

THE object of this paper is to indicate a method of deducing directly, from the stress equations and compatibility relations, the stresses in a heavy circular disk supported in a vertical plane by external forces. The problem of a circular disk placed vertically on a horizontal plane was solved by Michell (1900), and the distribution of stresses in a similar disk supported at the centre was obtained by Ghosh (1936 *a*), who also considered the case of a circular disk resting on two smooth pegs (1936 *b*).

\* Communicated by the Author.

By using bipolar co-ordinates and Airy's stress function, Mindlin (1938) determined the stresses in a heavy circular disk suspended vertically from an eccentric peg. Though these problems can be solved by the direct method outlined in this paper, they are not discussed here. As an illustration, only the problem of a heavy circular disk supported in a vertical plane by two symmetrically situated vertical forces and two horizontal forces acting on the rim, is considered. This includes, as particular cases, the problems of a disk supported vertically on two symmetrically placed pegs, rough or smooth. The method followed here was successfully employed in solving the boundary value problems of disks under no body forces in a paper accepted for publication elsewhere.

### 1. Method of Solution.

If the problem is considered to be one of plane stress in the vertical plane, then with the axis of  $x$  drawn vertically upwards and the axis of  $y$  in the horizontal direction, we have the stress equations of equilibrium

$$\left. \begin{aligned} \frac{\partial \widehat{xx}}{\partial x} + \frac{\partial \widehat{xy}}{\partial y} - w &= 0, \\ \frac{\partial \widehat{xy}}{\partial x} + \frac{\partial \widehat{yy}}{\partial y} &= 0 \end{aligned} \right\}, \quad \dots \dots \dots (1.1)$$

where  $w$  = weight per unit area, the thickness of the plate being unity.

The consistency relations are

$$\left. \begin{aligned} \nabla_1^2 \widehat{xx} + \frac{\partial^2 \Theta}{\partial x^2} &= 0, \\ \nabla_1^2 \widehat{xy} + \frac{\partial^2 \Theta}{\partial x \partial y} &= 0, \\ \nabla_1^2 \widehat{yy} + \frac{\partial^2 \Theta}{\partial y^2} &= 0, \\ \nabla_1^2 \Theta &= 0 \end{aligned} \right\}, \quad \dots \dots \dots (1.2)$$

where

$$\left. \begin{aligned} \Theta &= \widehat{xx} + \widehat{yy}, \\ \nabla_1^2 &= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \end{aligned} \right\}, \quad \dots \dots \dots (1.3)$$

From the first two equations of (1.2) we get

$$(x \nabla_1^2 \widehat{xx} + y \nabla_1^2 \widehat{xy}) + \left( x \frac{\partial^2 \Theta}{\partial x^2} + y \frac{\partial^2 \Theta}{\partial x \partial y} \right) = 0, \quad \left. \dots \dots \dots (1.4) \right\}$$

$$\text{or} \quad \nabla_1^2 (x \widehat{xx} + y \widehat{yy}) + r \frac{\partial}{\partial r} \left( \frac{\partial \Theta}{\partial x} \right) - 2 \left( \frac{\partial \widehat{xx}}{\partial x} + \frac{\partial \widehat{xy}}{\partial y} \right) = 0$$

where  $r^2 = x^2 + y^2$ .

Similarly from the second and the third of the equations (1.2) we obtain

$$\nabla_1^2 (x \widehat{xy} + y \widehat{yy}) + r \frac{\partial}{\partial r} \left( \frac{\partial \Theta}{\partial y} \right) - 2 \left( \frac{\partial \widehat{xy}}{\partial x} + \frac{\partial \widehat{yy}}{\partial y} \right) = 0. \quad \dots \dots (1.5)$$

Putting now

$$\left. \begin{aligned} r\widehat{rx} &= x\widehat{xx} + y\widehat{xy}, \\ r\widehat{ry} &= x\widehat{xy} + y\widehat{yy} \end{aligned} \right\}, \quad \dots \quad (1.6)$$

we get with the help of the equations (1.1)

$$\left. \begin{aligned} \nabla_1^2(r\widehat{rx} - \tfrac{1}{2}wr^2) + \frac{\partial}{\partial x} \left( r \frac{\partial \Theta}{\partial r} - \Theta \right) &= 0, \\ \nabla_1^2(r\widehat{ry}) + \frac{\partial}{\partial y} \left( r \frac{\partial \Theta}{\partial r} - \Theta \right) &= 0 \end{aligned} \right\} \quad \dots \quad (1.7)$$

Since  $\Theta$  is a plane harmonic function we can put

$$\Theta = \operatorname{Re} f(z), \quad \dots \quad (1.8)$$

where  $z$  stands for  $x+iy$ ,  $i=\sqrt{-1}$  and  $\operatorname{Re}$  denotes the real part of the complex function.

To find the solution of the equations (1.7), let us put

$$\left. \begin{aligned} r\widehat{rx} - \tfrac{1}{2}wr^2 &= \operatorname{Re} \left[ \frac{r^2 - a^2}{4} \phi(z) + a\mathcal{L}(z) \right], \\ r\widehat{ry} &= \operatorname{Re} \left[ \frac{r^2 - a^2}{4} i\psi(z) + a\mathcal{M}(z) \right] \end{aligned} \right\}, \quad \dots \quad (1.9)$$

where  $\phi(z)$ ,  $\psi(z)$ ,  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  are function of  $z$ .

The functions  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  are such that

$$\left. \begin{aligned} [r\widehat{rx}]_{r=a} - \tfrac{1}{2}wa^2 &= [\operatorname{Re} a\mathcal{L}(z)]_{r=a}, \\ [r\widehat{ry}]_{r=a} &= [\operatorname{Re} a\mathcal{M}(z)]_{r=a} \end{aligned} \right\} \quad \dots \quad (1.10)$$

On substituting the results (1.9) in (1.7), we find that these equations are satisfied if we put

$$z\phi(z) = z\psi(z) = f(z) - zf'(z).$$

Hence, as the solution of (1.7), we can write

$$\left. \begin{aligned} r\widehat{rx} - \tfrac{1}{2}wr^2 &= \operatorname{Re} \left[ \frac{r^2 - a^2}{4} \left\{ \frac{f(z) - zf'(z)}{z} \right\} + a\mathcal{L}(z) \right], \\ r\widehat{ry} &= \operatorname{Re} \left[ \frac{r^2 - a^2}{4} i \left\{ \frac{f(z) - zf'(z)}{z} \right\} + a\mathcal{M}(z) \right] \end{aligned} \right\} \quad \dots \quad (1.11)$$

Since

$$\Theta = \frac{\partial}{\partial x} (r\widehat{rx} - \tfrac{1}{2}wr^2) + \frac{\partial}{\partial y} (r\widehat{ry}),$$

we get

$$\operatorname{Re}[f(z) + zf'(z)] = 2a \operatorname{Re}[\mathcal{L}'(z) + i\mathcal{M}'(z)],$$

which is satisfied if we take

$$f(z) = 2a \frac{\mathcal{L}(z) + i\mathcal{M}(z)}{z}, \quad \dots \quad (1.12)$$

except when  $f(z) \propto z^{-1}$ . In this exceptional case a separate treatment is necessary, as shown in a previous paper (Sen, 1938). Thus with values of  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  determined from the boundary values of  $r\widehat{rx}$  and  $r\widehat{ry}$  as given in (1.10),  $f(z)$  can be found out from (1.12), and hence the stresses at any point of the plate can be obtained with the help of (1.11).

If a force acts at any point on the rim of the circular disk, we first write down the stresses  $\widehat{xx}_1$ ,  $\widehat{xy}_1$ ,  $\widehat{yy}_1$ , due to a force acting at the point on the *hypothesis* that the force acts on the edge of a semi infinite plate extending



indefinitely on the side of the disk. These results are well known. Then calculating  $r\widehat{r}x_1$ ,  $r\widehat{r}y_1$  from these stresses, with the help of (1.6) we can find their values on the edge of the circular disk. These stresses on the boundary  $r=a$  must be annulled by a second system of stresses given by  $r\widehat{r}x_2$ ,  $r\widehat{r}y_2$ , such that

$$\left. \begin{aligned} [r\widehat{r}x_2]_{r=a} &= -[r\widehat{r}x_1]_{r=a}, \\ [r\widehat{r}y_2]_{r=a} &= -[r\widehat{r}y_1]_{r=a} \end{aligned} \right\} \quad (1.13)$$

The boundary values of  $\widehat{r}x_2$  and  $\widehat{r}y_2$  being known, we can choose  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  to satisfy (1.10). In these expressions as well as in subsequent equations we put subscript 2 after  $\widehat{r}x$  and  $\widehat{r}y$ . In selecting the proper form for  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  we shall guard against the introduction of any singularity in the expressions for stresses. By proper choice of these two expressions we get  $f(z)$  from (1.12), and hence obtain the stresses  $r\widehat{r}x_2$ ,  $r\widehat{r}y_2$  at any point with the help of (1.11). The resultant stresses at any point are given by

$$\left. \begin{aligned} r\widehat{r}x &= r\widehat{r}x_1 + r\widehat{r}x_2, \\ r\widehat{r}y &= r\widehat{r}y_1 + r\widehat{r}y_2. \end{aligned} \right\} \quad (1.14)$$

In the case of a force acting inside the circular plate the procedure is the same except that in this case we write down  $\widehat{x}x_1$ ,  $\widehat{y}y_1$  on the *hypothesis* that the plate extends indefinitely in all directions from the point of application of the force.

## 2. Stresses in a heavy disk held in a vertical position by symmetrically situated forces acting on the rim.

Let two vertical forces each of magnitude P act at the points

$$A(-a \cos \alpha, -a \sin \alpha) \quad \text{and} \quad B(-a \cos \alpha, +a \sin \alpha).$$

Suppose also that there are two equal and opposite horizontal forces Q at these points. This will cover both the cases of smooth and rough supports at A and B. Stresses produced by such forces acting on the boundary of a semi-infinite plate are (cf. Timoshenko, 1932)

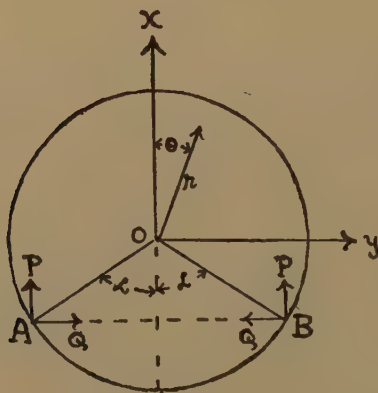
$$\left. \begin{aligned} \widehat{x}x_1 &= -\frac{2P}{\pi} \left[ \frac{(x+a \cos \alpha)^3}{r_1^4} + \frac{(x+a \cos \alpha)^3}{r_2^4} \right] \\ &\quad - \frac{2Q}{\pi} \left[ \frac{(x+a \cos \alpha)^2(y+a \sin \alpha)}{r_1^4} - \frac{(x+a \cos \alpha)^2(y-a \sin \alpha)}{r_2^4} \right], \\ \widehat{x}y_1 &= -\frac{2P}{\pi} \left[ \frac{(x+a \cos \alpha)^2(y+a \sin \alpha)}{r_1^4} + \frac{(x+a \cos \alpha)^2(y-a \sin \alpha)}{r_2^4} \right] \\ &\quad - \frac{2Q}{\pi} \left[ \frac{(x+a \cos \alpha)(y+a \sin \alpha)^2}{r_1^4} - \frac{(x+a \cos \alpha)(y-a \sin \alpha)^2}{r_2^4} \right], \\ \widehat{y}y_1 &= -\frac{2P}{\pi} \left[ \frac{(y+a \sin \alpha)^2(x+a \cos \alpha)}{r_1^4} + \frac{(y-a \sin \alpha)^2(x+a \cos \alpha)}{r_2^4} \right] \\ &\quad - \frac{2Q}{\pi} \left[ \frac{(y+a \sin \alpha)^3}{r_1^4} - \frac{(y-a \sin \alpha)^3}{r_2^4} \right] \end{aligned} \right\} \quad (2.1)$$

where

$$\left. \begin{aligned} r_1^2 &= (x+a \cos \alpha)^2 + (y+a \sin \alpha)^2, \\ r_2^2 &= (x+ \cos \alpha)^2 + (y-a \sin \alpha)^2 \end{aligned} \right\} \quad (2.2)$$

From these results we get

$$\left. \begin{aligned} r\widehat{r}x_1 &= -\frac{2P}{\pi} \left[ \frac{(x+a \cos \alpha)^2(r^2+ax \cos \alpha+ay \sin \alpha)}{r_1^4} \right. \\ &\quad \left. + \frac{(x+a \cos \alpha)^2(r^2+ax \cos \alpha-ay \sin \alpha)}{r_2^4} \right] \\ &\quad - \frac{2Q}{\pi} \left[ \frac{(x+a \cos \alpha)(y+a \sin \alpha)(r^2+ax \cos \alpha+ay \sin \alpha)}{r_1^4} \right. \\ &\quad \left. - \frac{(x+a \cos \alpha)(y-a \sin \alpha)(r^2+ax \cos \alpha-ay \sin \alpha)}{r_2^4} \right] \\ r\widehat{r}y_1 &= -\frac{2P}{\pi} \left[ \frac{(x+a \cos \alpha)(y+a \sin \alpha)(r^2+ax \cos \alpha+ay \sin \alpha)}{r_1^4} \right. \\ &\quad \left. + \frac{(x+a \cos \alpha)(y-a \sin \alpha)(r^2+ax \cos \alpha-ay \sin \alpha)}{r_2^4} \right] \\ &\quad - \frac{2Q}{\pi} \left[ \frac{(y+a \sin \alpha)^2(r^2+ax \cos \alpha+ay \sin \alpha)}{r_1^4} \right. \\ &\quad \left. - \frac{(y-a \sin \alpha)^2(r^2+ax \cos \alpha-ay \sin \alpha)}{r_2^4} \right] \end{aligned} \right\} \quad (2.3)$$



On the boundary  $r=a$  we have  $x=a \cos \theta$ ,  $y=a \sin \theta$ ,

$$r_1^2 = 2a^2[1 + \cos(\theta - \alpha)] \quad \text{and} \quad r_2^2 = 2a^2[1 + \cos(\theta + \alpha)].$$

Hence

$$\left. \begin{aligned} r\widehat{r}x_1 &= -\frac{P}{2\pi} (\cos \theta + \cos \alpha)^2 \left[ \frac{1}{1 + \cos(\theta - \alpha)} + \frac{1}{1 + \cos(\theta + \alpha)} \right] \\ &\quad - \frac{Q}{2\pi} (\cos \theta + \cos \alpha) \left[ \frac{\sin \theta + \sin \alpha}{1 + \cos(\theta - \alpha)} - \frac{\sin \theta - \sin \alpha}{1 + \cos(\theta + \alpha)} \right] \\ &= -\frac{P}{\pi} (1 + \cos \theta \cos \alpha) - \frac{Q}{\pi} \cos \theta \sin \alpha, \end{aligned} \right\} \quad (2.4)$$

$$\begin{aligned}
 r\widehat{ry}_1 = & -\frac{P}{2\pi}(\cos\theta + \cos\alpha) \left[ \frac{\sin\theta + \sin\alpha}{1 + \cos(\theta - \alpha)} + \frac{\sin\theta - \sin\alpha}{1 + \cos(\theta + \alpha)} \right] \\
 & - \frac{Q}{2\pi} \left[ \frac{(\sin\theta + \sin\alpha)^2}{1 + \cos(\theta - \alpha)} - \frac{(\sin\theta - \sin\alpha)^2}{1 + \cos(\theta + \alpha)} \right] \\
 = & -\frac{P}{\pi} \sin\theta \cos\alpha - \frac{Q}{\pi} \sin\theta \sin\alpha. \quad \dots \dots \dots (2.5)
 \end{aligned}$$

The stresses to be added for nullifying the above components on the boundary must be given by

$$\left. \begin{aligned}
 [r\widehat{rx}_2]_{r=a} = -[r\widehat{rx}_1]_{r=a} &= \frac{P}{\pi}(1 + \cos\theta \cos\alpha) + \frac{Q}{\pi} \cos\theta \sin\alpha, \\
 [r\widehat{ry}_2]_{r=a} = -[r\widehat{ry}_1]_{r=a} &= \frac{P}{\pi} \sin\theta \cos\alpha + \frac{Q}{\pi} \sin\theta \sin\alpha
 \end{aligned} \right\} \dots (2.6)$$

For equilibrium we shall have  $2P = \pi a^2 w$ . Therefore,

$$\left. \begin{aligned}
 [r\widehat{rx}_2]_{r=a} - \frac{1}{2}wa^2 &= \frac{wa^2}{2} \cos\theta \cos\alpha + \frac{Q}{\pi} \cos\theta \sin\alpha, \\
 [r\widehat{ry}_2]_{r=a} &= \frac{wa^2}{2} \sin\theta \cos\alpha + \frac{Q}{\pi} \sin\theta \sin\alpha
 \end{aligned} \right\} \dots \dots (2.7)$$

Then the functions  $a\mathcal{L}(z)$  and  $a\mathcal{M}(z)$  as given in (1.10) can be written as

$$\left. \begin{aligned}
 a\mathcal{L}(z) &= \left( \frac{wa^2}{4} \cos\alpha + \frac{Q}{2\pi} \sin\alpha \right) \left( \frac{z}{a} + \frac{a}{z} \right), \\
 a\mathcal{M}(z) &= -i \left( \frac{wa^2}{4} \cos\alpha + \frac{Q}{2\pi} \sin\alpha \right) \left( \frac{z}{a} - \frac{a}{z} \right)
 \end{aligned} \right\} \dots \dots (2.8)$$

Using the result (1.18) we obtain

$$f(z) = 2a \frac{\mathcal{L}(z) + \mathcal{M}(z)}{z} = wa \cos\alpha + \frac{2Q}{\pi a} \sin\alpha. \quad \dots \dots (2.9)$$

Substituting these values of  $f(z)$ ,  $\mathcal{L}(z)$  and  $\mathcal{M}(z)$  in (1.11) we get the expressions for  $r\widehat{rx}_2$  and  $r\widehat{ry}_2$  which together with  $r\widehat{rx}_1$  and  $r\widehat{ry}_1$  already obtained in (2.3) give the stresses at any point. From (2.1) we have

$$\widehat{xx}_1 + \widehat{yy}_1 = -wa^2 \left[ \frac{x+a\cos\alpha}{r_1^2} + \frac{x+a\cos\alpha}{r_2^2} \right] - \frac{2Q}{\pi} \left[ \frac{y+a\sin\alpha}{r_1^2} - \frac{y-a\sin\alpha}{r_2^2} \right]. \quad \dots \dots (2.10)$$

Since  $\widehat{rr} = 0$  on  $r = a$ , we have

$$\begin{aligned}
 [\theta\theta]_{r=a} &= [r\widehat{r} + \widehat{\theta\theta}]_{r=a} = [\widehat{xx}_1 + \widehat{yy}_1]_{r=a} + [Ref(z)]_{r=a} \\
 &= -\frac{wa(1 + \cos\theta \cos\alpha)}{\cos\alpha + \cos\alpha} - \frac{2Q}{\pi a} \frac{\cos\theta \sin\alpha}{\cos\theta + \cos\alpha} + wa \cos\alpha + \frac{2Q}{\pi a} \sin\alpha \\
 &= -\frac{wa \sin^2\alpha}{\cos\theta + \cos\alpha} + \frac{2Q}{\pi a} \frac{\sin\alpha \cos\alpha}{\cos\theta + \cos\alpha} \dots \dots \dots (2.11)
 \end{aligned}$$



If the pegs at A and B be equally rough, we can put

$$Q = P \tan (\alpha - \lambda) = \frac{w\pi a^2}{2} \tan (\alpha - \lambda),$$

where  $\lambda$  = angle of friction. Then

$$[\widehat{\theta\theta}]_{r=a} = - \frac{wa \sin \alpha \sin \lambda \sec (\alpha - \lambda)}{\cos \theta + \cos \alpha} \dots \dots \dots (2.12)$$

#### References.

- Ghosh, S., Bull. Cal. Math. Soc. xxviii. p. 145 (1936 *a*) ; xxix. p. 213 (1936 *b*).  
 Michell, J. H., Proc. Lond. Math. Soc. xxxii. p. 35 (1900).  
 Mindlin, R. D., Journal of Applied Physics, ix. p. 714 (1938).  
 Sen, B., Phil. Mag. ser. 7, xxvi. p. 98 (1938) ; Proc. Roy. Soc. (A) Lond. (as yet unpublished) (1941).  
 Timoshenko, S. P., 'Theory of Elasticity,' art. 29.

Department of Mathematics,  
 Bengal Engineering College,  
 Sibpore, Bengal.

#### XI. Notice respecting New Book.

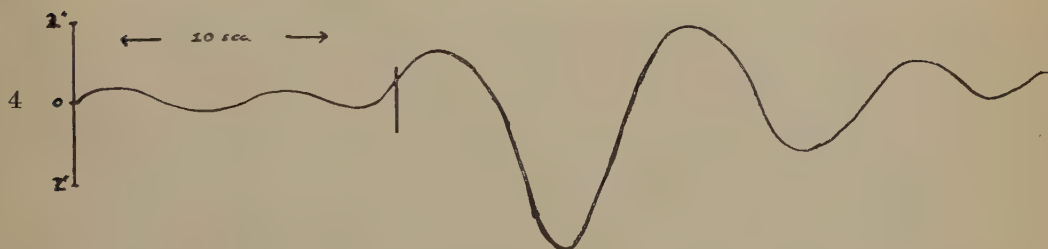
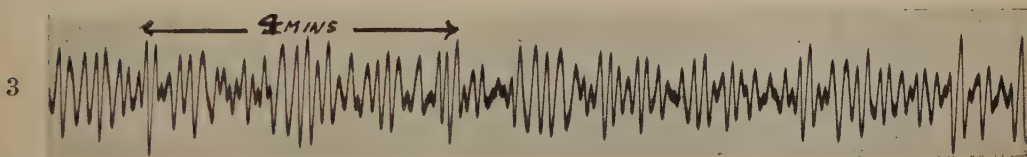
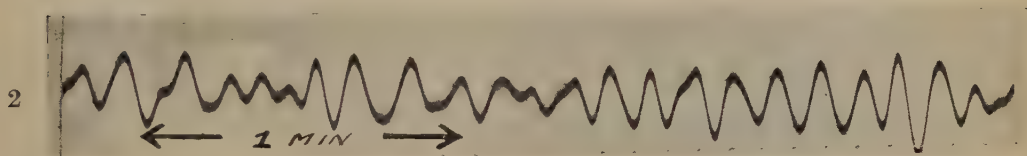
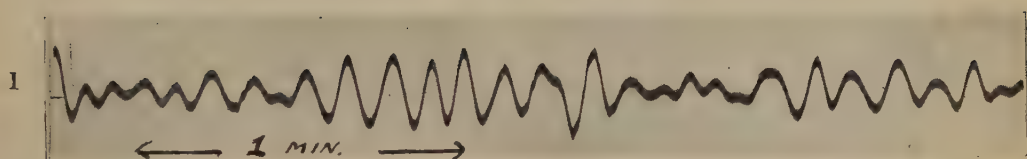
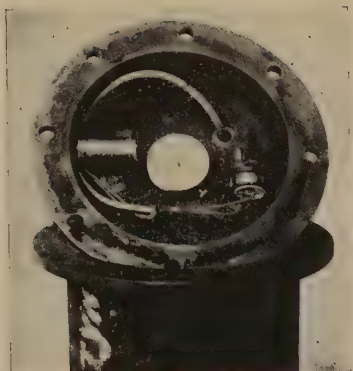
*The Advancement of Science*.—Vol. III. No. 12 (April 1946). [Pp. viii+92.]  
 (Published by the British Association for the Advancement of Science,  
 Burlington House, London, W. 1. Price 6s.)

WITH normal meetings rendered impossible, the British Association has carried on throughout the war by means of special conferences and by other methods, and one of its most useful activities has been to publish its new journal, 'The Advancement of Science.' The April number contains accounts of the four sessions of a conference on Scientific Research and Industrial Planning, which it would be quite impossible to review in any reasonable length. To select the views of one speaker, however, outstanding those views, would give a distorted picture of the general argument, and to give them all, even in the briefest space, would spoil them. All that can be said is that all those who are interested in, or concerned about, the future place of Science in our life, cannot afford to neglect the opportunity of reading the account here given.

Important as this conference is, it does not form the sole contents of the issue. In addition, there is an account of the Victory Congress of the French Association for the Advancement of Science, and a continuation of the articles on "Applications of Science," this time on the Place of Science in Town and Country Planning, from the pen of Prof. Dudley Stamp. Of more domestic interest is the report of the Council of the "British Ass." and the revised statutes and regulations. It is noted that the 1946 meeting will be a one-day gathering in London, on July 20, and that the 1947 and 1948 meetings will be held in Dundee and Brighton respectively.

J. H. A.

[The Editors do not hold themselves responsible for the views  
 expressed by their correspondents.]







XII. *Solar Radio Noise.*—I.

By Sir EDWARD APPLETON \*, F.R.S., and J. S. HEY †, M.Sc.‡

[Received June 28, 1946.]

[Plate II.]

1. *Introduction.*

IN two brief communications <sup>(1), (2)</sup> the authors have presented evidence indicating the occasional manifestation of electromagnetic radiation in the solar radio spectrum greatly in excess of the expected black-body intensity. Attention was first directed to this phenomenon by the reports of wireless amateurs § who, round about the period of the last sunspot maximum, noticed the existence of a curious hiss in their receivers on wave-lengths of about 10 metres, which they found to occur in day-time only, and to be associated with periods of solar activity. During the war, with the more sensitive and more directional ultra-short wave receivers in operation, many further reports of the reception of this noise have become available, while, during the recent period of pronounced sunspot activity (30th January to 14th February, 1946), the opportunity has been taken to make a specially detailed study of the phenomenon. In this communication we discuss the significance of all the evidence now available, and conclude that it confirms our theory that the noise is due directly to ultra-short wave electro-magnetic radiation, which originates in active areas on the sun, and travels from the sun to the earth with the speed of light. We also report that sudden large increases in intensity of solar noise have been observed to occur simultaneously with solar flares and radio fade-outs.

2. *Black-body Electro-magnetic Radiation from the Sun.*

For wave-lengths in the radio spectrum the flux of solar energy received at the earth's surface is given by  $F$  where

$$F = \frac{2\pi kT}{\lambda^2} \left( \frac{r}{R} \right)^2 \Delta f, \quad . . . . . (1)$$

where  $k$  is Boltzmann's constant,  $T$  the effective solar temperature,  $\lambda$  the wave-length,  $r$  the radius of the sun,  $R$  the earth-sun distance, and  $\Delta f$  the frequency-acceptance band. Now if we expose a half-wave

---

\* Department of Scientific and Industrial Research.

† Ministry of Supply.

‡ Communicated by the Authors.

§ Special mention should be made of the pioneer work of D. J. Heightman (Amateur Station G6DH) in this connection.

dipole with its length normal to the direction of the solar flux, the power absorbed by the dipole is proportional to  $\lambda^2$  ( $\Delta f$  being kept constant). Therefore, with radio aerials of this type, the solar noise received, due to the black-body radiation, is independent of wave-length. If the receiving installation has an aerial power gain  $G_r$  over a half-wave dipole, the available power  $p$  at the receiver may be shown to be given by

$$p = 0.41 kT \left( \frac{r}{R} \right)^2 G_r \Delta f. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now receiver noise may be expressed as being equivalent to a power  $p_r$  at the receiver terminals, given by

$$p_r = (N - 1) k T_0 \Delta f, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where  $T_0$  is room-temperature ( $290^\circ \text{K.}$ ), and  $N$  is the noise factor of the receiver. The practical values for  $N$ , range generally from about 3 to 8 in the metre wave-length range, and from 10 to 15 for centimetre wave-lengths. If we put  $N=10$  as a rough average, and assume that the sun radiates as a black body at temperature  $T=6000^\circ \text{K.}$ , the ratio of solar noise to receiver noise is given by

$$\frac{p}{p_r} = 2 \times 10^{-5} G_r. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Black-body noise power from the sun is therefore only about  $10^{-4}$  of receiver noise in the case of aerials with very low power gain (*e. g.*,  $G_r$  from 2 to 4). When, however, extremely short waves are used it is possible to employ directive aerial systems (*e. g.*, parabolic mirrors) of high power gain  $G_r$ , and so increase the ratio  $p/p_r$  proportionally. It is in this way that Reber<sup>(3)</sup> and Southworth<sup>(4)</sup> have succeeded in measuring solar radiation in the short radio wave end of the spectrum, and have shown that, at wave-lengths of 187 cm. and 10 cm. respectively, its magnitude conforms approximately to that emitted by a black body at a temperature of  $6000^\circ \text{K.}$

The evidence we present below shows, however, that in periods of sunspot activity, solar radio noise can be detected on ordinary aerials of low power gain, indicating the abnormal emission of energy very greatly in excess of that expected from a black body at  $6000^\circ \text{K.}$  On 8th February, 1946, for example, the value of solar noise measured in South-East England on 4.7 metres during the daytime was  $10^5$  to  $10^6$  times black-body value.

### 3. The Earlier Experimental Evidence.

As mentioned above, much evidence accumulated from amateur reports during the sunspot maximum period 1936-7, which indicated the correlation of the hiss phenomenon with solar activity, but quantitative data on the precise nature of the correlation have been lacking.

It is clear, for example, from an examination of equation (4), that unless very special steps are taken to increase the value of aerial power gain  $G_r$ , small enhancements of solar noise would not be detected. In the

case of the amateur wave-length of 10 metres it is not easy to increase  $G$  beyond a small number without the use of aerials of prohibitive size. Thus the early 1936-7 reports must have been related to enhancements of the order of  $10^4$  and above, and not to lower values. But enough information was available to indicate that a strong enhancement of solar noise was often followed by a sudden ionospheric disturbance and that, during that disturbance, the noise usually disappeared.

During the war it has been possible to study the phenomenon, using receiving installations with a much greater value of  $G_r$ . Moreover, since such systems were equipped with accurate direction finding facilities, it has also been possible to demonstrate that the radio noise comes from the direction of the sun. Evidence of this kind was obtained, using Army radar receiving equipment on 27th and 28th February, 1942, on the 4 to 6 metre band <sup>(2)</sup>.

The radar sets which observed the solar noise on this occasion were situated in widely separated parts of Great Britain. The operators determined the bearing according to the normal practice for finding the direction of a source of interference. It was found that the bearings moved throughout the day, and were always within a few degrees of that of the sun. The most striking results came from two sites about 150 miles apart, where the elevation was also measured. Observers on these sites were able to follow the noise source continuously in bearing and elevation, and observation through the equipment telescope confirmed that their equipment was directed at the sun. As the aperture of the telescope was  $\pm 5^\circ$ , we may conclude that the mean deviation in both bearing and elevation was not more than about two degrees, which is within the expected accuracy of the equipment for measurements on a noise signal. The noise was not observed at night at any station.

These results show that the noise must have been caused by the direct propagation of electromagnetic noise radiations from the sun. The possibility that the noise might have originated from violent perturbations of the ionization in the ionosphere is excluded, since, in such a case, the source would have been most concentrated at the sub-solar point, and not situated in the direction of the sun itself. Moreover, a diffuse source, such as an intensely ionized layer, would not have given accurate bearing and elevation measurements.

Supporting evidence that the bearing was the same as that of the sun was provided on the same occasion by the reports of several R.A.F. radar stations operating on 7 metres and 13 metres. These stations reported that their first observation of increased noise was on the 23rd February, and it increased in intensity until about 28th February. The intensity was negligible on and after the 1st March.

Details of solar activity for the period in question have been given by Newton <sup>(5)</sup> and Stratton <sup>(6)</sup>. The sunspot crossed the solar disk in heliographic latitude  $7^\circ$  north during the period 22nd February to 7th March. Central meridian passage occurred on 28th February. The maximum area of the sunspot was 2000 millionths of the sun's hemisphere



and, as Newton has pointed out, sunspots greater than 1500 millionths are of special interest since, in seven such cases out of ten, a magnetic storm occurs near the time of the central meridian passage. In this instance a very extensive and brilliant solar flare occurred from approximately 12.00 hours to 15.30 hours, on 28th February, and the accompanying radio fade-out, which began at 12.00 hours, lasted on some circuits until 20.00 hours. At Abinger a crochet of exceptional range was recorded on the magnetograms, beginning at 12.01 with partial recovery about an hour later. A great magnetic storm broke out with extreme suddenness at 07.27 on the 1st March.

The days of the highest intensity of the solar noise, 27th and 28th February, 1942, occurred when the sunspot was near the central meridian passage, and at a time of intense sunspot activity, several solar flares being observed during this period in addition to the very brilliant one to which reference has just been made. The sunspot decreased in size after 28th February, at the same time as the solar noise subsided.

#### 4. *The Recent Experimental Evidence.*

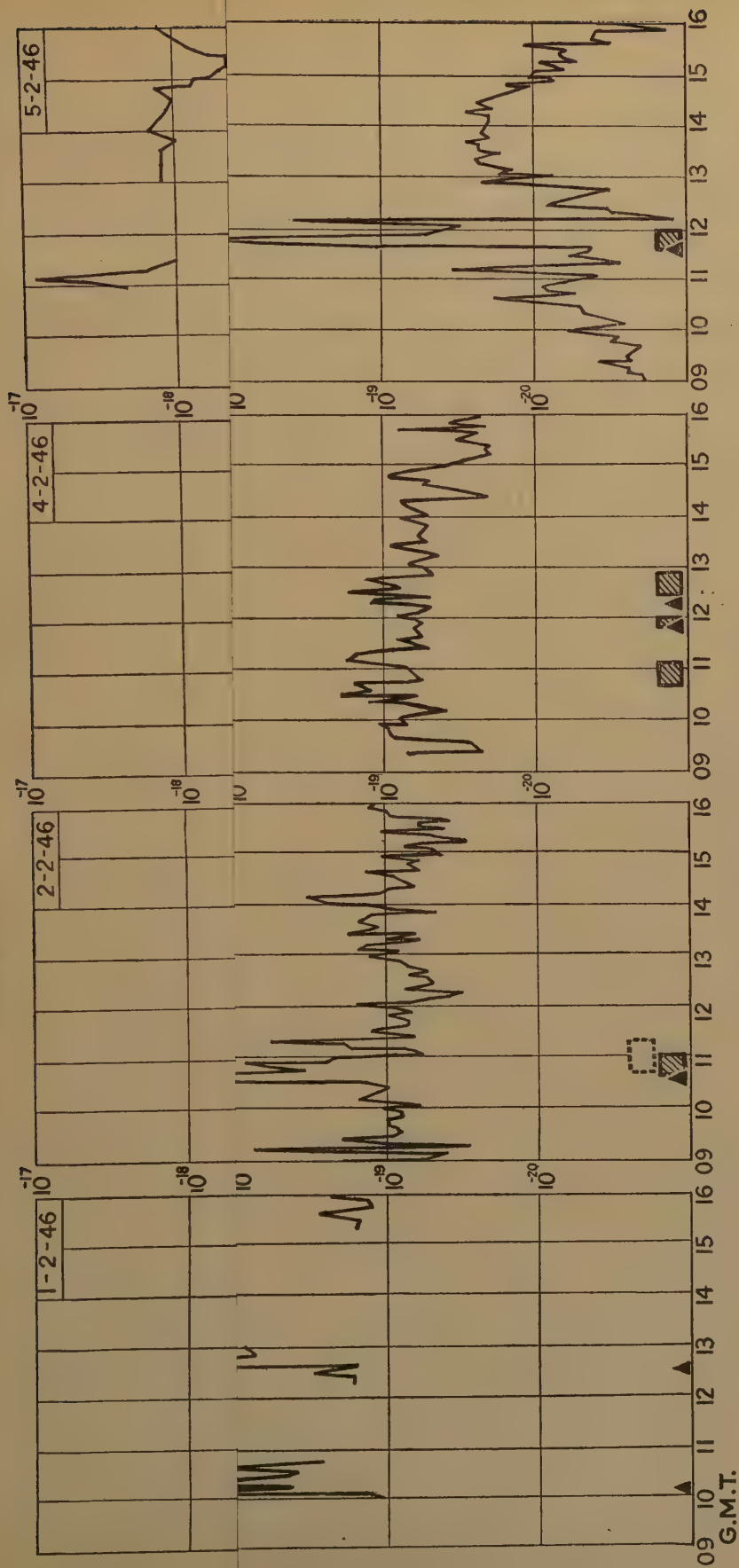
The recent manifestation of sunspot activity during the period 30th January to 14th February, 1946, provided an excellent opportunity for further examination of solar noise, and in particular for studying more closely its connection with other solar and terrestrial phenomena. This has been made possible by the practically continuous record of solar noise power flux on a wave-length of 4.7 metres, which was made at the research station of the Operational Research Group, Ministry of Supply \*. Using a directional aerial system for which  $G_r$  was approximately 100, the ratio of theoretical black-body radiation to receiver noise,  $p/p_r$  (see equation (1)) was  $2 \times 10^{-3}$ . As the recording meter was sufficiently sensitive to indicate variations of 5 per cent. of receiver noise, it was possible to detect all occurrences of solar noise for which the enhancement over the black-body value was greater than about 100. It was found that the abnormal solar noise emission was at an extremely high level throughout the whole of the passage of the great sunspot over the sun's visible hemisphere during the period 30th January to 14th February, 1946. This spot attained a maximum area of about 5000 millionths of the sun's hemisphere, and therefore considerably exceeded in size the one observed during February 1942. A photograph of the sunspot is shown in fig. 1 (Pl. II.) by courtesy of the Astronomer Royal.

The solar noise varied somewhat irregularly, the general level being highest on the 5th and 8th February when, on a wave-length of 4.7 metres, it was equal to approximately 10,000 times receiver noise level, but occasional peaks reaching 100,000 times noise level were recorded. With the exception of those days when the record is very incomplete, the solar noise observations of solar noise power flux at 4.7 metres for horizontal polarization are shown, for 09.00 to 16.00 hours each day, in fig. 2.

---

\* The authors are specially indebted to S. J. Parsons, J. W. Phillips and G. S. Stewart for these observations.

Fig. 2.



Solar noise power flux at 4.7 metres.

In watts/sq. metre/C.P.S. bandwidth (horizontal polarization). Plotted as ordinate with time in hours G.M.T. as abscissa.

▨ Cable and wireless fade-out. ▲ Onset of fade-out reported by M. Bureau. □ Solar flare.

Correlation of solar noise enhancement, radio fade-out and solar flare can be seen on 2nd, 9th and 11th February. Simultaneous solar noise enhancement and radio fade-out also evident on 10th and 13th February, solar flare data not yet available.





### 5. *The Correlation of Solar Noise and other Phenomena.*

It has been known since the period of the last sunspot maximum of 1936-7 that there is a close correlation between bright eruptions (solar flares) and sudden ionospheric disturbances. It is believed that, simultaneously with the onset of many solar flares, there is emitted, from the sun, a form of ionizing radiation which enhances the ionization normally existing in the D Layer. It is this enhancement of D Layer ionization which causes the heavy absorption of short radio waves on long-distance radio circuits. The heavy absorption effects disappear as soon as the solar flare abates, due to the rapid recombination and attachment process for electrons which operate at these relatively low ionospheric levels. Very often the fade-out effect which is noted on short wave radio circuits traversing the sun-lit hemisphere, is accompanied by an enhancement of signal on long wave-circuits. Such enhancements have been made the subject of special study in the case of long wave atmospheric disturbances by R. Bureau. It is clear that the temporarily-enhanced D Layer ionization in such cases :—

- (a) Increases the reflection coefficients of the D Layer for long waves which are reflected at its lower boundary.
- (b) Increases the absorption of short waves which pass through the D Layer on their way to and from the E or F Layers, at which levels they are reflected.

Prompted by the above considerations we have attempted to compare the recent measurements of solar noise with all the relevant solar and terrestrial data for the period in question. For that purpose we have examined :—

- (a) Information on radio fade-outs as supplied to us by Cable and Wireless, Ltd.
- (b) Information on fade-outs and simultaneous enhancement of long wave atmospherics as notified by R. Bureau <sup>(7)</sup>.
- (c) Observations on solar flares, sunspot areas and magnetic data, supplied by the Astronomer Royal.
- (d) Ionospheric data supplied by the Radio Research Station, Slough, and by the Inter-Services Ionospheric Bureau, Great Baddow.

The transit of the sunspots across the sun's disk during the period 30th January to 14th February, 1946, was accompanied by the usual terrestrial events which the investigation of the last ten years has led us to expect. The sunspot reached central meridian position on 5/6th February and, a day or two later, a large magnetic storm was recorded at the Royal Observatory Magnetic Station at Abinger. This storm began abruptly at 10h. 20m. G.M.T., on 7th February, and lasted for about 36 hours. It was accompanied by a display of the northern lights which, while visible in Scotland, was obscured in the southern part of England by cloud. The ranges of the magnetic elements were as high as  $1.3^\circ$  in declination, more than 500  $\gamma$  in horizontal force, and nearly 400  $\gamma$  in

vertical force. The storm was remarkable more for the agitation of the magnetic traces rather than for the ranges, which have been exceeded on seven or eight other occasions during the last 11-year solar cycle, 1934–1944. At the onset of the storm the sunspot group was about 1.9 days ( $=25^\circ$  of solar longitude) past the central meridian.

During the period covering the solar noise record, shown in fig. 2, Cable and Wireless, Ltd., reported 10 fade-outs. The most striking result which emerges from a comparison of the times of the fade-outs and the

Date.	Time of onset of fade-out reported by Cable and Wireless (duration in brackets)	Time of fade-out and long wave enhancement of atmospheres reported by M. Bureau.	Time of onset of solar noise increase.	Solar flares observed at the Greenwich Royal Observatory.
2.2.46	09.12 (1.8 hrs.)	09.10 *	Gradual increase from 09.10. Max. value at 10.15	09.54–11.26. Intensity 3. Max. at 10.13
9.2.46	11.53 (0.1 hrs.)	11.50	Sharp rise from 11.52 to peak at 11.54	11.50–12.27. Intensity 3. Max. at 11.56
10.2.46	—	10.08	Sharp rise at 10.08	No data available
10.2.46	—	12.30	Sharp rise between 12.30 and 12.34	No data available
11.2.46	10.35 (0.4 hrs.)	10.32	Sharp rise at 10.35	10.40–11.15. Intensity 3
13.2.46	11.35 (0.4 hrs.)	11.35	Sharp rise between 11.35 and 11.38	No data available

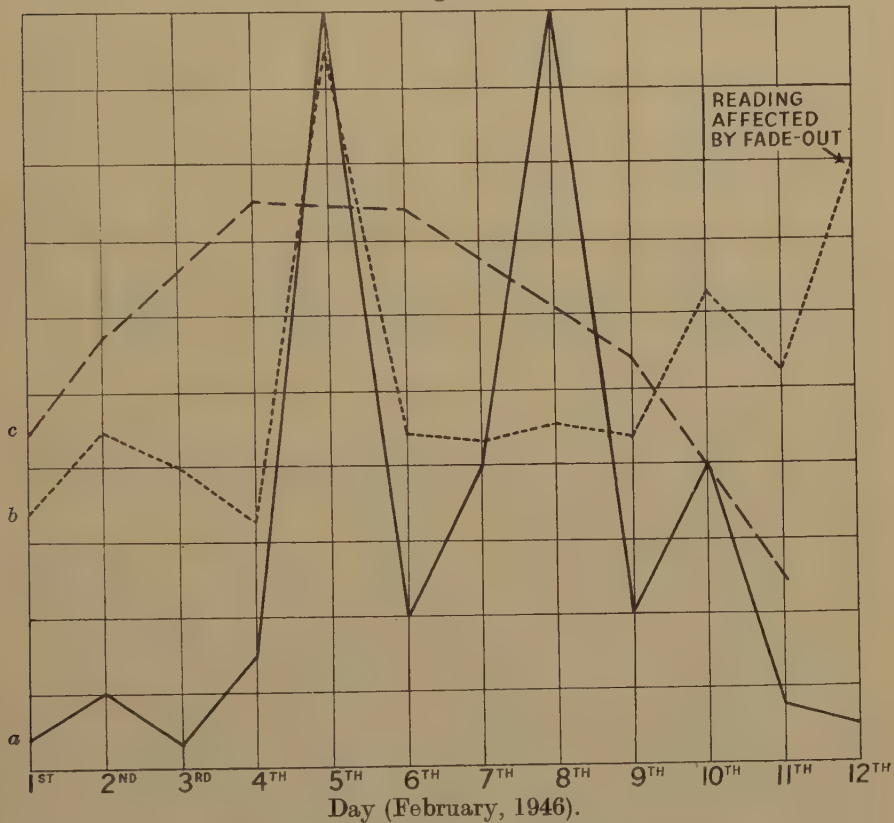
data in fig. 2 is the simultaneity between four major fade-outs and four very marked enhancements of solar noise. Moreover, of four simultaneous fade-outs and long-wave enhancements of atmospheres reported by M. Bureau, additional to the anomalies listed by Cable and Wireless, two corresponded to marked enhancements of solar noise. Further, solar flares of intensity 3 were noted at Greenwich on each of three of the above occasions, the only ones for which the weather was suitable for spectrohelioscopic observation in this country. These simultaneous occurrences, which are tabulated below, are also indicated in fig. 1 (Pl. II.).

\* No atmospheric enhancement was noted on this occasion.

We may therefore take as a definite result emerging from these comparisons that, when a solar flare occurs which is accompanied by a radio fade-out, there is also in many cases a marked enhancement of solar noise. One further correlation is worthy of note, namely a magnetic crochet of exceptional size reported by J. Coulomb and E. Selzer <sup>(8)</sup> on 13th February at 11.35, which coincided with the fade-out and solar noise peak.

Now in studies of solar-terrestrial relationships in the past, the simultaneity of the solar flare and the enhancement of D Layer ionization has been used to prove that the agency responsible for the augmentation

Fig. 3.



Comparison of relative magnitudes of

- (a) Average level of solar radio noise —————
- (b) "D" layer absorption index at noon - - - - -
- (c) Ratio of projected sunspot area to that of visible disk — — — — —

of ionization travels from the sun to the earth with the velocity of light. Since the radio noise has been found to increase at the same time as these phenomena develop, we may therefore regard it as virtually certain that the agency causing solar noise also travels to the earth, along with the ionizing and visible radiations, with the speed of light.

In addition to the more detailed examination described above, we have



thought it of interest to compare the daily average level of solar radio noise with ionospheric absorption and sunspot area. For that purpose there is plotted, in fig. 3, (a) the daily average level of solar radio noise, (b) the measured values of D Layer absorption at noon, and (c) the ratio of the projected sunspot area to that of the visible disk. It will be seen that the solar radio noise reached a first maximum value when the sunspot was central and then, after a diminution, recovered to approximately the same peak value again about three days later. It will also be noted that, while there is not a very intimate correlation, the general level of D Layer ionization (fade-out periods excepted) was raised during the period when solar noise was marked. The curves of both solar radio noise and D Layer absorption show an asymmetry with respect to that of sunspot projected area. On the return of the sunspot, after half a solar rotation, at the end of February, the area was only 1200 millionths of the sun's hemisphere, and the solar noise, though still detectable, had weakened very considerably in intensity.

#### 6. *Directional Observations on Solar Radio Noise.*

During the recent manifestations of solar noise a number of Government research establishments co-operated in making observations on the direction of arrival of solar radio noise, and also on the magnitude of the received radio flux on different wave-lengths. These establishments may be listed as follows :—

- (a) Operational Research Group, Ministry of Supply, working on 0.1, 1.5, 4.7, 8, 12 and 15 metres.
- (b) Telecommunications Research Establishment, using equipments on 0.1, 0.5 and 1.5 metres wave-length.
- (c) 60 Group, R.A.F., where investigations on radar sets on 0.1, 1.5 and 10 to 15 metres wave-length were organized by E. Eastwood, F. Copsey, and H. Crossland.
- (d) H.M.S. 'Collingwood,' where A.C. Copisarow obtained results on 0.1, 1.5, 3.25 and 7.5 metres wave-length.
- (e) Admiralty Signal Establishment Extension, where H.M. 'Bristow' made observations on 1.5, 3.3 and 7.5 metres wave-length.

The results of all these establishments agree in indicating that the bearing of the source was that of the sun within the expected limits of accuracy of the equipments used, most of which on this occasion depended on measurements of signal strength with a beam that was fixed in elevation. For a noise signal of fluctuating amplitude the accuracy was generally of the order of a few degrees. The results also showed that the source effectively traversed the vertical polar diagram of the equipment in accordance with the supposition that the radiation originated in the sun. Several observers noted that the signal was detectable a few minutes before sunrise and after sunset. This anomaly is possibly explicable in terms of ionospheric and atmospheric refraction, and of diffraction at the earth's surface.

On only one occasion was an additional natural source of noise radiation suspected. A.C. Copisarow, of H.M.S. 'Collingwood,' reported that during the afternoon of 7th February, a pronounced source of noise also occurred in the region of  $350^\circ$  bearing, approximately in the direction of the Magnetic North Pole. This effect was observed at two stations nine miles apart on 1.4 metres; and Copisarow has suggested that these noise radiations may be connected with a magnetic storm which commenced at 10.20 on that day. No confirmatory evidence of the phenomenon has been available from other sources, though it should be added that no other establishment appears to have made a search for alternative sources at that particular time. (Cosmic, or galactic, radio noise, was, of course, present throughout the whole period of the observation, but its magnitude was negligible in comparison with the solar radio noise observed.)

### 7. *The Observed Spectrum of Solar Radio Noise.*

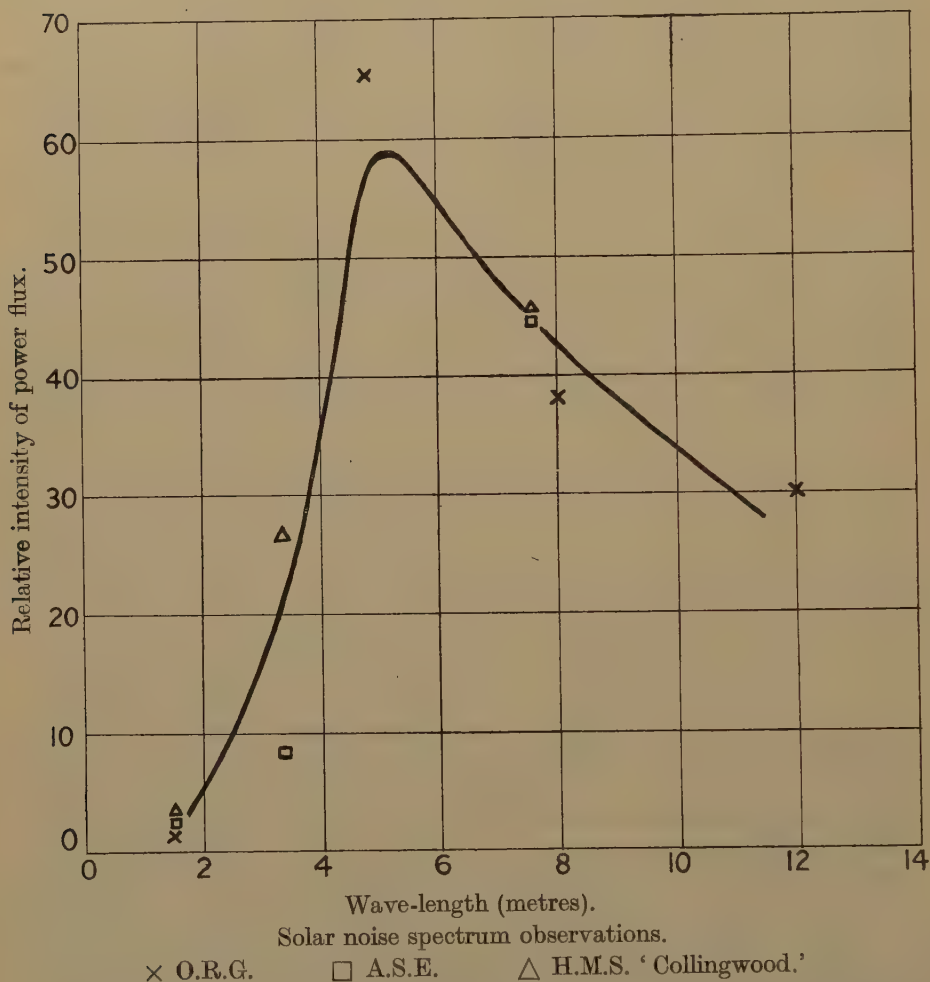
Special interest is attached to the spectrum of solar radio noise. For that reason we have collected all the information available, from the observing stations mentioned above, which bears on the relations between solar radio noise and the wave-length of observation. From the negative results obtained by several investigators, it is clear that no special enhancement of solar radiation has been found during the recent period of sunspot activity on wave-lengths as short as 0.1 metre. The power gain of the equipment employed on this wave-length was considerably greater than that of the metre wave-length equipments, although the effect of this was partially offset by the higher noise factor. Similarly, observations by the Telecommunications Research Establishment on 0.5 metre yielded negative results. Indeed, the increase of radio noise associated with sunspot activity only becomes significant, without very special techniques, when the wave-length is increased to about  $1\frac{1}{2}$  metres. This result is in accordance with the recent observations of Pawsey, Payne-Scott and McCready<sup>(9)</sup>. With increases of wave-length up to 5 metres, there is, however, a very steep rise in the magnitude of solar noise. The exact shape of the spectral curve beyond this wave-length is somewhat uncertain, because of the disturbing influence of ionospheric effects to which we refer below.

In fig. 4 the relative intensities of solar radio power flux, measured on the earth's surface by three groups of investigators, are plotted against wave-length. The graph was obtained first by plotting the results obtained by the Operational Research Group, those for wave-lengths of 8 and 12 metres being the average of comparisons made near noon on three days. When a rough mean curve was drawn through these points, the relative power flux values, measured by the Admiralty Signal Establishment and H.M.S. 'Collingwood,' were normalized to the curve at the  $7\frac{1}{2}$  metre wave-length, and the values at other wave-lengths were found to be in general agreement with the curve.

In order to illustrate the quantitative relation of solar radio noise to

the black-body radiation expected for a solar temperature of  $6000^{\circ}\text{K.}$ , we have plotted on the same diagram (a) the solar black-body spectrum, and (b) the spectrum of solar radio noise corresponding to a power flux of  $10^{-18}$  watts per square metre per cycle per second for a wave-length of 4.7 metres. This relationship is shown in fig. 5, and illustrates the marked excess of solar radio noise intensity over the expected black-body value.

Fig. 4.

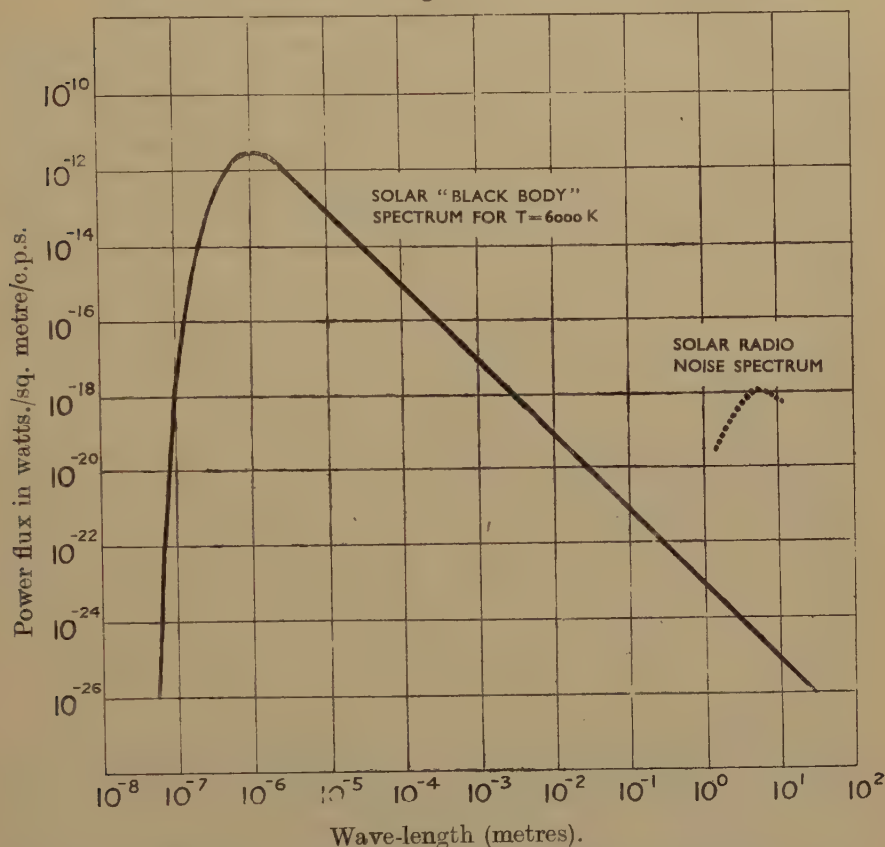


It should, however, be emphasized that the spectrum of solar radio noise observed at ground-level is not the same as that which would be measured at a point immediately outside the earth's atmosphere. Indeed, it is clear that, due to ionospheric reflection, there is always a critical "cut-off" wave-length depending on the existing characteristics of the ionosphere; and radiation of wave-length longer than this critical wave-length is prevented from penetrating the ionosphere and reaching



ground-level. Moreover, radiation of wave-length just less than the critical value is subject both to absorptive attenuation and to spatial attenuation or reinforcement which are markedly dependent on wave-length. We believe that such differential effects have been operative in determining the shape of the spectrum curve in figs. 4 and 5 for the range of wave-lengths exceeding about 5 metres. Support for this view is the fact that marked variations in the ratio of long wave flux to short wave

Fig. 5.



Comparison of solar radio noise spectrum and theoretical solar black body spectrum.

flux have been noticed at different times of the day, the long wave flux being relatively greater in the early morning than at mid-day.

The problem of the influence of atmospheric and ionospheric attenuation on radio noise of extra-terrestrial origin is evidently a general and important one, for it relates to both galactic and solar noise\*. We shall discuss it in greater detail in Part II. of this paper.

\* In some more recent experiments (July 27th and 28th, 1946) we found that solar radio noise, on that occasion, was circularly polarized.

*Acknowledgments.*

We wish to acknowledge the interest and support of Sir John Lennard-Jones, F.R.S., D.G.S.R. (Defence), and of Lt.-Col. H. A. Sargeant, Superintendent, Operational Research Group, Ministry of Supply. We also wish to express our indebtedness to the Astronomer Royal for the supply of solar and other data.

The paper is published by permission of the Department of Scientific and Industrial Research, and the Ministry of Supply.

*Summary.*

An account is given of experiments carried out, during a period of sunspot activity, on the recently discovered phenomenon of solar radio noise. Measurements of the spectrum of such noise, as observed at ground-level, have been made, though it is pointed out that the shape of the spectral curve on the long wave-length side is deformed by ionospheric influences. Large and sudden increases in intensity of solar radio noise have been observed to occur simultaneously with solar flares and short wave radio fade-outs.

*References.*

- (1) Appleton, 'Nature,' clvi. p. 534 (1945).
- (2) Hey, 'Nature,' clvii. p. 47 (1946).
- (3) Reber, *Astroph. Journ.* c. p. 279 (1944).
- (4) Southworth, *Journ. Franklin Inst.* cccxxix. p. 285 (1945).
- (5) Newton, 'The Observatory,' lxiv. p. 260 (1942).
- (6) Stratton, 'Nature,' clvii. p. 47 (1946).
- (7) Bureau, *Comptes Rendus*, cccxii. p. 597 (1946).
- (8) Coulomb and Selzer, *Comptes Rendus*, cccxii. p. 599 (1946).
- (9) Pawsey, Payne-Scott and McCready, 'Nature,' clvii. p. 158 (1946).

XIII. *An Expression for the Sine of a Fourier Series.*

By C. STRACHEY and P. J. WALLIS\*.

[Received June 29, 1945.]

IN many branches of applied mathematics it is very useful to represent a periodic function as a Fourier series. It has often been necessary to find the Fourier coefficients (or at least some of them) of the sine of a Fourier series, but the formula for these coefficients does not seem to have been published before, although it is quite easy to obtain.

The Bessel functions  $J_r(x)$  of integral order  $r$  are sometimes defined by the generating function

$$e^{\frac{1}{2}x(t-1/t)} = \sum_{r=-\infty}^{\infty} t^r J_r(x),$$

\* Communicated by the Authors.

which, for  $t=e^{i(r\theta+\delta_r)}$  and  $x=c_r$ , becomes

$$e^{ic_r \sin(r\theta+\delta_r)} = \sum_{q=-\infty}^{\infty} e^{iq(r\theta+\delta_r)} J_q(c_r).$$

This is a well-known result and this note generalises it to include more than one harmonic.

Representing the first  $n$  harmonics of the given function as

$$\begin{aligned} \phi_n &= \sum_{r=1}^n c_r \sin(r\theta+\delta_r), \\ e^{i\phi_n} &= \prod_{r=1}^n e^{ic_r \sin(r\theta+\delta_r)} = \prod_{r=1}^n \left\{ \sum_{q_r=-\infty}^{\infty} e^{iq_r \delta_r} J_{q_r}(c_r) \cdot e^{ir q_r \theta} \right\} \\ &= \sum_{q_n=-\infty}^{\infty} \sum_{q_{n-1}=-\infty}^{\infty} \dots \sum_{q_1=-\infty}^{\infty} e^{i \sum_{r=1}^n q_r \delta_r} \left\{ \prod_{r=1}^n J_{q_r}(c_r) \right\} e^{i \sum_{r=1}^n r q_r \theta}. \end{aligned}$$

If, in addition,  $p$  and  $c_0$  are given,

$$e^{i(p\theta+c_0+\phi_n)} = \sum_{(q)} \left[ e^{i(c_0+\sum_{r=1}^n q_r \delta_r)} \left\{ \prod_{r=1}^n J_{q_r}(c_r) \right\} e^{i(p+\sum_{r=1}^n r q_r)\theta} \right],$$

where the summation is taken over all positive and negative integral values of the  $n$   $q$ 's.

If the Fourier series is written in the form

$$e^{i(p\theta+c_0+\phi_n)} = \sum_{m=-\infty}^{\infty} \gamma_m e^{im\theta},$$

where the  $\gamma$ 's may be complex, then the above expansion immediately gives any particular  $\gamma_m$  by restricting the summation over the  $q$ 's such that  $p + \sum_{r=1}^n r q_r = m$ . Then

$$\begin{aligned} \gamma_m &= \sum_{(q)} \left[ e^{i(c_0+\sum_{r=1}^n q_r \delta_r)} \left\{ \prod_{r=1}^n J_{q_r}(c_r) \right\} \right] \\ \sum_{r=1}^n r q_r &= m-p. \end{aligned}$$

The above expression is exact, but its usefulness depends on being able to neglect all except the first few terms which have small  $q$ 's. In any particular case inspection will show which terms can be neglected. Quite often, however, it is known that the original Fourier coefficients decrease like powers,

$$c_r=O(x^x) \text{ with } 0 < x < 1.$$

In this case  $J_{q_r}(c_r)=O(x^{q_r})$  and  $\sum |r q_r|$  can be used as a measure of the size of the term in the infinite series.

The following table (for  $n=4$  and  $p < 4$ ) shows the leading terms in the expansion.

Using this table it is easy to see which terms occur for any given  $p$  simply by running the eye down the right column and noting the values of the  $q$ 's in any row having the right value for  $m$ . The table can be extended quite easily if needed.



As an example, taking  $n=4$ ,  $p=2$  and  $m=-1$ ,

$$e^{i(2\theta + c_0 + c_1 \sin \theta + \delta_1 + c_2 \sin 2\theta + \delta_2 + c_3 \sin 3\theta + \delta_3 + c_4 \sin 4\theta + \delta_4)} = \sum_{-\infty}^{\infty} \gamma_m e^{im\theta},$$

where

$$\begin{aligned} \gamma_{-1} = & -J_0(c_1)J_0(c_2)J_1(c_3)J_0(c_4)e^{i(c_0-\delta_3)} + J_1(c_1)J_1(c_2)J_0(c_3)J_0(c_4)e^{i(c_0-\delta_1-\delta_2)} \\ & - J_3(c_1)J_0(c_2)J_0(c_3)J_0(c_4)e^{i(c_0-3\delta_1)} + \text{higher terms.} \end{aligned}$$

$\Sigma  rq_r $	$q_1$	$q_2$	$q_3$	$q_4$	Values of $m=p+\Sigma rq_r$ for $p=$									
					-4	-3	-2	-1	0	1	2	3	4	
0	0	0	0	0	-4	-3	-2	-1	0	1	2	3	4	
1	1	0	0	0	-3	-2	-1	0	1	2	3	4	5	
	-1	0	0	0	-5	-4	-3	-2	-1	0	1	2	3	
2	2	0	0	0	-2	-1	0	1	2	3	4	5	6	
	0	1	0	0										
	0	-1	0	0	-6	-5	-4	-3	-2	-1	0	1	2	
	-2	0	0	0										
3	3	0	0	0	-1	0	1	2	3	4	5	6	7	
	1	1	0	0										
	0	0	1	0	-3	-2	-1	0	1	2	3	4	5	
	-1	1	0	0	-5	-4	-3	-2	-1	0	1	2	3	
	1	-1	0	0										
	0	0	-1	0	-7	-6	-5	-4	-3	-2	-1	0	1	
	-1	-1	0	0										
	-3	0	0	0										
4	4	0	0	0	0	1	2	3	4	5	6	7	8	
	2	1	0	0										
	1	0	1	0	-2	-1	0	1	2	3	4	5	6	
	0	2	0	0	-4	-3	-2	-1	0	1	2	3	4	
	0	0	0	1	-6	-5	-4	-3	-2	-1	0	1	2	
	-1	0	1	0										
	2	-1	0	0	-8	-7	-6	-5	-4	-3	-2	-1	0	
	-2	1	0	0										
	1	0	-1	0										
	0	0	0	-1										
	0	-2	0	0										
	-1	0	-1	0										
	-2	-1	0	0										
	-4	0	0	0										

Usually it is required to find a particular (say the  $m$ th) coefficient of the sine or cosine of  $(p\theta + c_0 + \phi_n)$ , and this follows immediately by taking the imaginary or real part of  $\gamma_m e^{im\theta} + \gamma_{-m} e^{-im\theta}$ .

The authors' thanks are due to Standard Telephones & Cables, Ltd., for permission to publish the above paper.

# XIV. *Hahn's Functions* $S_m(\alpha)$ and $U_m(\alpha)$ .

By C. STRACHEY and P. J. WALLIS \*.

[Received June 29, 1945.]

## 1. *Summary and Introduction.*

IN a paper <sup>(1)</sup> on the calculation of fields in certain resonators, Hahn introduced two new functions :

$$\begin{aligned} -S_m(\alpha) &= \sum_{n=1}^{\infty} \frac{m^2 \sin^2 n\pi\alpha}{n(m^2 - n^2\alpha^2)} \\ \text{and} \quad U_m(\alpha) &= \sum_{n=1}^{\infty} \frac{m^2 n \sin^2 n\pi\alpha}{\alpha^2 \left( n^2 - \frac{m^2}{\alpha^2} \right)^2} \quad \text{with } 0 < \alpha < 1 \end{aligned}$$

and used these functions to shorten his calculations. Since this time, Hahn's method has been used for certain similarly-shaped resonators and Hahn's two functions usually help to shorten the solution considerably. Hahn himself only gave a small table of  $S_m(\alpha)$  and a few values of  $U_m(\alpha)$ .

In this report closed expressions are derived for the case of  $\alpha$  rational, and are used to produce a much more comprehensive table of  $S_m(\alpha)$  and a slightly smaller table of  $\frac{1}{m} U_m(\alpha)$ . In a concluding section integral expressions, power series in  $\alpha$ , and asymptotic series in  $m$  are given which together facilitate the calculation for values of  $\alpha$  not given in the tables.

## 2. *Closed Expressions for Rational $\alpha$ .*

For many purposes it is only necessary to know the values of Hahn's functions for rational  $\alpha (=p/q, \text{ say})$ , and in this case it is comparatively easy to obtain closed expressions for the functions which include those previously given by Goddard <sup>(2)</sup> as a special case (when  $p=1$ ). For  $\alpha=p/q$

$$\begin{aligned} -S_m(\alpha) &= \sum_{n=1}^{\infty} \frac{m^2 \sin^2 n\pi\alpha}{n(m^2 - n^2\alpha^2)} \\ &= \sum_{n=1}^{\infty} \frac{m^2 \sin^2 \frac{np}{q} \pi}{n \left( m^2 - \frac{n^2 p^2}{q^2} \right)} \\ &= \sum_{k=1}^{q-1} \frac{\sin^2 \frac{kp}{q} \pi}{2q} \sum_{r=0}^{\infty} \left( \frac{2}{r + \frac{k}{q}} - \frac{1}{r + \frac{k}{q} + \frac{m}{p}} - \frac{1}{r + \frac{k}{q} - \frac{m}{p}} \right) \end{aligned}$$

by writing  $n=rq+k$ ,

$$= \frac{1}{2q} \sum_{k=1}^{q-1} \left[ \psi \left( \frac{k}{q} + \frac{m}{p} \right) + \psi \left( \frac{k}{q} - \frac{m}{p} \right) - 2\psi \left( \frac{k}{q} \right) \right] \sin^2 k\pi\alpha, \quad (2.1)$$

---

\* Communicated by the Authors.

since 
$$\sum_0^{\infty} \left( \frac{1}{n+a} - \frac{1}{n+b} \right) = \psi(b) - \psi(a),$$

where  $\psi(x)$  is the digamma function tabulated in the 'B.A. Mathematical Tables,' Vol. I., and in Davis, 'Tables of Higher Mathematical Functions.'

It can readily be seen that when  $m/p$  is an integer, then the term from the original series with  $n=qm/p$  vanishes in the limit, and consequently the above formula is valid even in this case.

For  $p=1$

$$\begin{aligned} -S_m \left( \frac{1}{q} \right) &= \frac{1}{2q} \sum_{k=1}^{q-1} \sum_{r=0}^{m-1} \frac{2}{\frac{k}{q} + r} \sin^2 k\pi\alpha \\ &= \sum_{n=1}^{mq-1} \frac{\sin^2 \frac{n\pi}{q}}{n}, \end{aligned}$$

agreeing with Goddard's results.

Similarly,

$$\begin{aligned} \frac{1}{m} U_m(\alpha) &= \sum_{n=1}^{\infty} \frac{mn \sin^2 n\pi\alpha}{\alpha^2 \left( n^2 - \frac{m^2}{\alpha^2} \right)^2} \\ &= \sum_{k=1}^{q-1} \frac{\sin^2 \frac{kp\pi}{q}}{4pq} \sum_{r=0}^{\infty} \left( \frac{1}{\left( r + \frac{k}{q} - \frac{m}{p} \right)^2} - \frac{1}{\left( r + \frac{k}{q} + \frac{m}{p} \right)^2} \right), \end{aligned}$$

when  $m/\alpha$  is not an integer,

$$= \frac{1}{4pq} \sum_{k=1}^{q-1} \left[ \psi' \left( \frac{k}{q} - \frac{m}{p} \right) - \psi' \left( \frac{k}{q} + \frac{m}{p} \right) \right] \sin^2 k\pi\alpha. \quad (2.2)$$

If  $m/p$  is an integer,  $m/p=s$  say, then the term for  $n=qs$  is exceptional and gives a contribution

$$\frac{\pi^2\alpha}{4},$$

obtained as the limit for  $\alpha$  tending to  $p/q$ .

For  $p=1$

$$\begin{aligned} \frac{1}{m} U_m(\alpha) &= \frac{\pi^2}{4q} + \frac{1}{4q} \sum_{k=1}^{q-1} \left[ \psi' \left( \frac{k}{q} - m \right) - \psi' \left( \frac{k}{q} + m \right) \right] \sin^2 k\pi\alpha \\ &= \frac{\pi^2}{4q} + \frac{1}{4q} \sum_{k=1}^{q-1} \sum_{r=0}^{m-1} \frac{2 \sin^2 k\pi\alpha}{\left( \frac{k}{q} + r \right)^2} \\ &= \frac{\pi^2}{4q} + \frac{1}{2} \sum_{n=1}^{mq-1} \frac{\sin^2 \frac{n\pi}{q}}{n^2}, \end{aligned}$$

agreeing with Goddard.



The two formulæ (2.1) and (2.2) simplify for particular values of  $p/q$  and the following results can be easily obtained :

$$\left. \begin{aligned} -S_m\left(\frac{1}{2}\right) &= \psi(2m) - \frac{1}{2}\psi(m) + \frac{1}{2}\gamma, \text{ where } \gamma \text{ is Euler's constant,} \\ \frac{1}{m}U_m\left(\frac{1}{2}\right) &= \frac{\pi^2}{4} - \psi'(2m) + \frac{1}{4}\psi'(m) \\ -S_m\left(\frac{1}{3}\right) &= \frac{3}{4}\psi(3m) - \frac{1}{4}\psi(m) + \frac{1}{2}\gamma \\ \frac{1}{m}U_m\left(\frac{1}{3}\right) &= \frac{\pi^2}{4} - \frac{9}{8}\psi'(3m) + \frac{1}{8}\psi'(m) \\ -S_m\left(\frac{2}{3}\right) &= \frac{3}{4}\psi\left(\frac{3m}{2}\right) - \frac{1}{4}\psi\left(\frac{m}{2}\right) + \frac{1}{2}\gamma \\ \frac{1}{m}U_m\left(\frac{2}{3}\right) &= \frac{\pi^2}{4} - \frac{9}{16}\psi'\left(\frac{3m}{2}\right) + \frac{1}{16}\psi'\left(\frac{m}{2}\right) \\ -S_m\left(\frac{1}{4}\right) &= \frac{1}{2}\psi(4m) + \frac{1}{4}\psi(2m) - \frac{1}{4}\psi(m) + \frac{1}{2}\gamma \\ \frac{1}{m}U_m\left(\frac{1}{4}\right) &= \frac{\pi^2}{4} - \psi'(4m) + \frac{1}{4}\psi'(2m) + \frac{1}{8}\psi'(m) \\ -S_m\left(\frac{3}{4}\right) &= \frac{1}{2}\psi\left(\frac{4m}{3}\right) + \frac{1}{4}\psi\left(\frac{2m}{3}\right) - \frac{1}{4}\psi\left(\frac{m}{3}\right) + \frac{1}{2}\gamma \\ \frac{1}{m}U_m\left(\frac{3}{4}\right) &= \frac{\pi^2}{4} - \frac{1}{3}\psi'\left(\frac{4m}{3}\right) - \frac{1}{12}\psi'\left(\frac{2m}{3}\right) + \frac{1}{24}\psi'\left(\frac{m}{3}\right) \end{aligned} \right\} \quad (2.3)$$

The above formulæ only represent  $-S_m(\alpha)$  and  $\frac{1}{m}U_m(\alpha)$  for integral  $m$  but the  $\psi$ 's and  $\psi$ 's are defined for all  $m$ , and for non-integral  $m$  the formulæ given represent different functions from those given by the original Hahn series, only agreeing with them for integral  $m$ . The original functions satisfy the relation

$$\frac{1}{m}U_m(\alpha) = -\frac{1}{2}\frac{\partial}{\partial m}[-S_m(\alpha)],$$

whereas the above interpolatory functions satisfy

$$\frac{1}{m}U_m(\alpha) = \frac{\pi^2}{4} - \frac{1}{2}\frac{\partial}{\partial m}[-S_m(\alpha)].$$

### 3. Tables of $-S_m(\alpha)$ and $\frac{1}{m}U(\alpha)$ .

The formulæ of paragraph 2 are quite sufficient to calculate the values of  $-S_m(\alpha)$  and  $\frac{1}{m}U_m(\alpha)$  for integral  $m$  and rational  $\alpha$ . It is very easy to see that

$$\left. \begin{aligned} S_{hp}\left(\frac{p}{p+q}\right) &= S_{hq}\left(\frac{q}{p+q}\right) \\ \text{and } U_{hp}\left(\frac{p}{p+q}\right) &= U_{hq}\left(\frac{q}{p+q}\right) + \frac{\pi^2}{4}h(p-q) \end{aligned} \right\}, \quad (3.1)$$

which serve as useful checks on calculated values. The following results have been obtained :

$$-S_m(\alpha).$$

$\frac{m}{a}$	1	2	3	4	5	6	7	8	9	10
0	1.21883	1.55718	1.75824	1.90148	2.01277	2.10378	2.18076	2.24746	2.30631	2.35897
0.1	1.21058	1.54893	1.74998	1.89323	2.00452	2.09553	2.17251	2.23921	2.29806	2.35071
0.2	1.18562	1.52387	1.72491	1.86814	1.97943	2.07044	2.14742	2.21412	2.27297	2.32562
0.3	1.14324	1.48104	1.68198	1.82518	1.93645	2.02745	2.10443	2.17113	2.22998	2.28263
0.4	1.08205	1.41855	1.61923	1.76234	1.87357	1.96454	2.04151	2.10820	2.16704	2.21968
0.5	1.00000	1.33333	1.53333	1.67619	1.78730	1.87821	1.95513	2.02180	2.08062	2.13326
0.6	0.89371	1.22019	1.41855	1.56078	1.67159	1.76234	1.83916	1.90576	1.96454	2.01714
0.7	0.75775	1.06998	1.26438	1.40505	1.51510	1.60538	1.68198	1.74836	1.80705	1.85957
0.8	0.58285	0.86499	1.04939	1.18562	1.29335	1.38231	1.45801	1.52387	1.58212	1.63435
0.9	0.35105	0.56366	0.71779	0.83861	0.93700	1.01956	1.09142	1.15546	1.21058	1.26240
1.0	0	0	0	0	0	0	0	0	0	0
0.25	1.16667	1.50476	1.70577	1.84900	1.96028	2.05128	2.12826	2.19497	2.25381	2.30646
0.75	0.67604	0.97604	1.16667	1.30571	1.41493	1.50476	1.58102	1.64725	1.70577	1.75818
$\frac{1}{3}$	1.12500	1.46250	1.66339	1.80657	1.91784	2.00883	2.08580	2.15250	2.21135	2.26400
$\frac{2}{3}$	0.80685	1.12500	1.32114	1.46250	1.57289	1.66339	1.74007	1.80657	1.86529	1.91784

$$\frac{1}{m} U_m(\alpha).$$

$\frac{m}{\alpha}$	1	2	3	4	5	6	7	8	9	10
0	2.22763	2.34388	2.38452	2.40509	2.41750	2.42579	2.43172	2.43618	2.43964	2.44241
0.25	2.22796	2.34393	2.38454	2.40510	2.41750	2.42579	2.43172	2.43618	2.43964	2.44241
0.50	2.23370	2.34481	2.38481	2.40522	2.41757	2.42583	2.43175	2.43619	2.43965	2.44242
0.75	2.26752	2.35248	2.38759	2.40650	2.41826	2.42624	2.43201	2.43638	2.43978	2.44251
1.0	2.46740	2.46740	2.46740	2.46740	2.46740	2.46740	2.46740	2.46740	2.46740	2.46740
$\frac{1}{3}$	2.22872	2.34403	2.38457	2.40511	2.41751	2.42580	2.43173	2.43618	2.43964	2.44241
$\frac{2}{3}$	2.25000	2.34806	2.38592	2.40572	2.41783	2.42598	2.43185	2.43626	2.43970	2.44246

No attempt is made in this report to give values for  $S_0(\alpha)$  since a good table of this function has been calculated by Lowan, and given in a paper by Whinnery and Jamieson<sup>(3)</sup>. Further, the function  $S_0(\alpha)$  is not the limit of  $S_m(\alpha)$  as  $m \rightarrow 0$  but is given by

$$S_0(\alpha) = \lim_{m \rightarrow 0} \frac{S_m(\alpha)}{m^2}.$$

Comparison of the above results with Hahn's original table<sup>(1)</sup> shows that he has several rounding-off errors, and had not used the check (3.1). The values for  $\alpha=0$  were obtained from formulæ given in the next section.

The above tables are not sufficient for interpolation with regard to  $\alpha$  in all regions to the full five decimals, but spot values can be calculated according to the formulæ given in paragraphs 2 and 4.

#### 4. Miscellaneous Formulæ.

##### (i) Integral Formulæ.

It can be proved that :

$$\sum_{n=1}^{\infty} \frac{\cos n\theta}{n+\alpha} = \cos \alpha\theta \left\{ \ln \frac{1}{2 \sin \theta/2} - \psi(\alpha+1) - \gamma + \frac{1}{\alpha} \sin^2 \frac{\alpha\theta}{2} \right. \\ \left. + \frac{1}{2} \int_0^{\theta} (1 - \cos \alpha\phi) \cot \frac{\phi}{2} d\phi \right\} + \frac{1}{2} \sin \alpha\theta \left\{ \pi - \int_0^{\theta} \sin \alpha\phi \cot \frac{\phi}{2} d\phi \right\}.$$

Using this result, a little reduction gives

$$-S_m(\alpha) = - \sum_{n=1}^{\infty} \frac{\frac{m^2}{\alpha^2} \sin^2 n\pi\alpha}{n \left( n^2 - \frac{m^2}{\alpha^2} \right)} \\ = \frac{1}{4} \sum_{n=1}^{\infty} (1 - \cos 2n\pi\alpha) \left( \frac{2}{n} - \frac{1}{n - \frac{m}{\alpha}} - \frac{1}{n + \frac{m}{\alpha}} \right) \\ = -\sin^2 m\pi \left[ \ln \frac{1}{2 \sin \pi\alpha} - \frac{1}{2} \psi \left( 1 + \frac{m}{\alpha} \right) - \frac{1}{2} \psi \left( 1 - \frac{m}{\alpha} \right) - \gamma \right] \\ + \alpha \int_0^{\pi} \frac{\sin m\phi \sin m(\phi - 2\pi) d\phi}{\tan \alpha\phi}.$$

For  $m$  integral this gives

$$-S_m(\alpha) = \alpha \int_0^{\pi} \frac{\sin^2 m\phi}{\tan \alpha\phi} d\phi. \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

Similarly

$$\frac{1}{m} U_m(\alpha) = \frac{\pi}{2} \sin 2m\pi \left[ \ln \frac{1}{2 \sin \pi\alpha} - \frac{1}{2} \psi \left( 1 + \frac{m}{\alpha} \right) - \frac{1}{2} \psi \left( 1 - \frac{m}{\alpha} \right) - \gamma \right] \\ - \frac{\sin^2 m\pi}{4\alpha} \left[ \psi' \left( 1 + \frac{m}{\alpha} \right) - \psi' \left( 1 - \frac{m}{\alpha} \right) \right] \\ - \frac{\alpha}{2} \int_0^{\pi} \{ (\pi - \phi) \sin 2m(\pi - \phi) - \pi \sin 2\pi m \} \frac{d\phi}{\tan \alpha\phi}$$



For  $m$  integral,  $m/\alpha$  non-integral :

$$\frac{1}{m} U_m(\alpha) = \frac{\alpha}{2} \int_0^\pi (\pi - \phi) \frac{\sin 2m\phi d\phi}{\tan \alpha\phi}. \quad \dots \quad (4.2)$$

For  $m$  and  $m/\alpha$  integral :

$$\frac{1}{m} U_m(\alpha) = \frac{\alpha}{2} \int_0^\pi (\pi - \phi) \frac{\sin 2m\phi d\phi}{\tan \alpha\phi} - \frac{\pi^2 \alpha}{4}. \quad \dots \quad (4.3)$$

(ii) *Power Series in  $\alpha$ .*

It is easy to obtain the expansions of  $S_m(\alpha)$  and  $U_m(\alpha)$  in powers of  $\alpha$  from the above integrals by expanding the  $\frac{1}{\tan \alpha\phi}$  in the integrand as a power series in  $\alpha\phi$  and evaluating the integrals term by term. Using the results

$$\int_0^\pi \phi^{2p-1} \cos 2m\phi d\phi = \pi^{2p} (2p-1)! \sum_{q=1}^{p-1} \frac{(-)^{q-1}}{(2p-2q)! (2m\pi)^{2q}}$$

$$\text{and } \int_0^\pi (\pi - \phi) \phi^{2p-1} \sin 2m\phi d\phi = \frac{\pi^{2p} (2p-1)!}{m} \sum_{q=1}^{p-1} \frac{(-)^q q}{(2p-2q)! (2m\pi)^{2q}},$$

then

$$\begin{aligned} -S_m(\alpha) &= C_s(2m\pi) - \sum_{p=1} \frac{2^{2p-1} B_p \alpha^{2p}}{2p!} \left[ \frac{\pi^{2p}}{2p} + \sum_{q=1}^{p-1} \frac{\pi^{2p} (2p-1)! (-)^{q-1}}{(2p-2q)! (2m\pi)^{2q}} \right] \\ &= C_s(2m\pi) - \frac{\alpha^2 \pi^2}{12} - \frac{\alpha^4 \pi^4}{360} \left( 1 + \frac{3}{m^2 \pi^2} \right) - \frac{\alpha^6 \pi^6}{945} \left( \frac{1}{6} + \frac{5}{4m^2 \pi^2} - \frac{15}{4m^4 \pi^4} \right), \\ &\quad \dots \quad (4.4) \end{aligned}$$

where

$$C_s(x) = \int_0^x \frac{1 - \cos \theta}{\theta} d\theta$$

and the  $B_p$ 's are the Bernoulli numbers

$$B_1 = \frac{1}{6}, \quad B_2 = \frac{1}{30}, \quad B_3 = \frac{1}{42}, \quad B_4 = \frac{1}{30}, \quad B_5 = \frac{5}{66}.$$

Similarly

$$\begin{aligned} U_m(\alpha) &= \frac{m\pi}{2} Si(2m\pi) + \sum_{p=1}^\infty \frac{B_p}{4p} (2\alpha\pi)^{2p} \sum_{q=1}^{p-1} \frac{(-)^{q-1} q}{(2p-2q)! (2m\pi)^{2q}} \\ &= \frac{m\pi}{2} Si(2m\pi) + \frac{\alpha^4 \pi^2}{120 m^2} + \frac{\alpha^6 \pi^4}{756 m^2} \left( 1 - \frac{6}{m^2 \pi^2} \right) \\ &\quad + \frac{\alpha^8 \pi^6}{5400 m^2} \left( 1 - \frac{15}{m^2 \pi^2} + \frac{135}{2m^4 \pi^4} \right), \quad \dots \quad (4.5) \end{aligned}$$

where

$$Si(x) = \int_0^x \frac{\sin \theta}{\theta} d\theta.$$

(iii) *Asymptotic Series in  $m$ .*

If the double sums in (4.4) and (4.5) are reversed, the terms are ordered

in powers of  $(1/m)$ . If at the same time  $C_s(2m\pi)$  and  $Si(2m\pi)$  are expanded, the following asymptotic expansions can be obtained :

$$-S_m(\alpha) = \frac{1}{2}\gamma + \frac{1}{2}\ln\left(\frac{2m}{\alpha}\sin\alpha\pi\right) + \frac{1}{2}\sum_{p=1}^{\infty}\frac{(-)^p B_p}{2p}\left(\frac{\alpha}{m}\right)^{2p} \\ + \frac{1}{2}\sum_{p=1}^{\infty}(-)^p\left(\frac{\alpha}{2m}\right)^{2p}D_{\alpha\pi}^{2p}\ln\sin x, \quad (4.6)$$

and

$$U_m(\alpha) = \frac{m\pi^2}{4} - \frac{1}{4} + \frac{1}{4}\sum_{p=1}^{\infty}(-)^p B_p\left(\frac{\alpha}{m}\right)^{2p} + \frac{1}{2}\sum_{p=1}^{\infty}(-)^p p\left(\frac{\alpha}{2m}\right)^{2p}D_{\alpha\pi}^{2p}\ln\sin x. \\ \dots (4.7)$$

Where  $D_{\alpha\pi}^{2p}\ln\sin x$  denotes the value at  $x=\alpha\pi$  of the  $2p$ 'th derivative of  $\ln\sin x$ . Writing  $s = \sin\alpha\pi$

$$D_{\alpha\pi}^2\ln\sin x = -\frac{1}{s^2},$$

$$D_{\alpha\pi}^4\ln\sin x = \frac{4}{s^2} - \frac{6}{s^4},$$

$$D_{\alpha\pi}^6\ln\sin x = -\frac{16}{s^2} + \frac{120}{s^4} - \frac{120}{s^6}.$$

The authors' thanks are due to Standard Telephones and Cables, Ltd. for permission to publish the above paper.

#### References.

- (1) W. C. Hahn, "A New Method for the Calculation of Cavity Resonators," J. A. P., xii. pp. 62-68, (1 Jan. 1941).
- (2) L. S. Goddard, "A Problem in the Summation of Series," Proc. Camb Phil. Soc., xxxix. Part 3, p. 200 (Oct. 1943).
- (3) J. R. Whinnery and H. W. Jamieson, "Equivalent Circuits for Discontinuities in Transmission Lines," P. I. R. E., xxxii. pp. 98-114 (2 Feb. 1944).

### XV. The Influence of Spray Particle Size and Distribution in the Combustion of Oil Droplets.

By R. P. PROBERT\*.

[Received November 12, 1945.]

#### SUMMARY.

The evaporation of liquid sprays is examined theoretically using the known law of size distribution in a spray and an experimental law concerning the effect of size on the evaporation of single droplets. The investigation is concerned not with the mechanism of evaporation, but

\* Communicated by the Author.

only with the effects of size and distribution. The mean diameter and rate of evaporation of given distributions are evaluated, and also the variation of the volume of a spray as evaporation proceeds, for all sprays likely to be met with in practice. The effect of distribution on combustion intensity is discussed, and the investigation shows what characteristics are desirable in a spray from the point of view of combustion. The data sheets provided should assist in correlating experimental evidence of the effects of atomiser characteristics on combustion losses and flame length, etc.

## 1. INTRODUCTION.

QUALITY of atomisation has always been recognized as one of the fundamental factors controlling the combustion of liquid fuel sprays and much attention has been given to methods of measuring particle sizes. A sound technique has been developed by the Asiatic Petroleum Co., Ltd., and results of the application of this technique are now available. It should now be possible for quantitative measurements to be taken of the effects of particle size and distribution in combustion, but there is an obvious need for a theoretical treatment whereby such measurements may be correlated. A first attempt at such a treatment is given here. It is based on the law of distribution in a spray of liquid amongst drops of different sizes, which is justified by the experimental work done on atomisers, and a simple law of evaporation which is known to apply to the slow evaporation of small drops, although there is not yet any work reported dealing with evaporation in conditions comparable to a combustion chamber. It seeks to combine these laws in order to demonstrate the effects of spray parameters on evaporation and hence on combustion.

## 2. THE DISTRIBUTION OF FUEL IN A SPRAY.

The sieving technique of measuring size distribution in a spray has been developed by the Asiatic Petroleum Co., Ltd.\* Practically all the results available show that the size distribution follows the Rosin-Rammler relation.

$$R = e^{-(x/\bar{x})^n}, \quad . . . . . (1)$$

where  $R$  = volume or weight fraction of the spray composed of drops greater in diameter than  $x$ .

$\bar{x}$  = constant (dimensions of diameter), which will be called the size constant.

$n$  = constant (number), which will be called the distribution constant.

The constants  $\bar{x}$  and  $n$  define any spray completely; particle size measurements on sprays from atomisers are always analysed and the values of  $\bar{x}$  and  $n$  deduced and quoted.

---

\* The paper was first issued as a Power Jets Report R. 1107 in July 1945.

## 3. EVAPORATION.

Hitherto, it has been assumed in work on combustion that a single drop evaporates, for otherwise constant conditions, at a rate proportional to its surface area. There is no experimental work on evaporation under conditions comparable to a combustion chamber to support this assumption. On the contrary, experimental work from other sources shows that small droplets do in fact evaporate at a rate proportional to the first power of the diameter. A few such investigations are reviewed below.

For the evaporation of a drop into still air, Maxwell\* showed that

$$E = \frac{4\pi r D}{K(p_1 - p_0)},$$

where  $E$  = rate of evaporation.

$r$  = radius of drop.

$p_1$  = saturation pressure at surface temperature.

$p_0$  = vapour pressure of surrounding air.

$D$  = diffusion constant.

$K$  = constant depending on air temperature and density.

Experiments 2, 3 on the evaporation of droplets in an air stream confirm the law of proportionality to diameter. For example, the evaporation of small water droplets of 0.2 mm. diameter and less obeyed the law,

$$\frac{dr^2}{dt} = -(A - B e^{-\alpha v}),$$

where  $A$ ,  $B$ ,  $\alpha$  are constants depending on temperature and humidity, and  $v$  = relative velocity of the wind.

In a combustion chamber, evaporation must be much faster than in the experiments referred to, but it is difficult to escape the conclusion that the drops will in any case evaporate at a rate proportional to diameter. The only positive surface effect in combustion is the reception of heat by radiation, but it is easy to show that radiation alone can account for only a small proportion of the heat required for the evaporation of a typical droplet in the time available.

It is assumed here that droplets evaporate at a rate proportional to diameter, not (diameter)<sup>2</sup> or surface area.

## 4. MEAN DIAMETER OF A SPRAY.

In any spray there will be drops of all sizes, and for such a spray it is possible to define several mean diameters, such as the numerical mean diameter and the weight mean diameter. But since the best spray is that which evaporates fastest, it is convenient to define a mean diameter representative of the overall rate of evaporation.

---

\* Research Engineer with Messrs. Power Jets (R. & D.) Ltd.



If we have a given volume of liquid consisting of drops of all sizes, that volume will evaporate fastest when the sum of the diameters of all the drops in the spray is greatest. Pictorially, we require that if all the drops in a spray are laid end to end, the length of the line of drops should be greatest for the fastest evaporation of the whole. On this property, it is possible to define a mean diameter which is the diameter of a droplet which evaporates at the same rate per unit volume as the aggregate rate per unit volume of liquid in the spray.

For the evaluation of this diameter, we shall consider the spray as consisting of several groups of droplets, each group consisting of a number of drops of roughly equal size.

Let  $d$ ,  $v$  be the diameter and volume of individual drops respectively.

Let  $v_n$  be a small volume fraction of the spray in which the drops are closely grouped about the diameter  $d_n$ .

Let  $d_m$  be the mean diameter.

Then 
$$\frac{d_m}{\pi/6 d_m^3} = \frac{\Sigma d}{\Sigma v}, \quad \dots \dots \dots (2)$$

and in the  $n$ th group of drops

$$\frac{\Sigma d}{\Sigma v} = \frac{d_n}{\pi/6 d_n^3}.$$

$$\therefore \quad \frac{\Sigma d}{n} = \frac{v n}{\pi/6 d_n^2} \quad \dots \dots \dots (3)$$

$$\therefore \quad d_m = \sqrt{\frac{\Sigma v}{\Sigma \frac{v_n}{d_n^2}}} \quad \dots \dots \dots (4)$$

or for unit volume of the spray.

$$d_m = \sqrt{\frac{1}{\Sigma \frac{v_n}{d_n^2}}} \quad \dots \dots \dots (5)$$

Provided the size distribution in a fuel spray is known, expression (3) allows the mean diameter to be calculated and hence the rate of evaporation of the whole spray.

## 5. MEAN DIAMETER OF THE INJECTED SPRAY.

The size distribution has been given in equation (1).

$$R = e^{-(x/\bar{x})^n}.$$

Taking differentials

$$\delta R = \frac{-n x^{n-1}}{\bar{x}^n} \cdot e^{-(x/\bar{x})^n} \delta x,$$

but  $\delta R$  and  $x$  correspond to  $v_n$  and  $d_n$  in the notation of (5).

$$\therefore d_m = \sqrt{\frac{1}{\Sigma \frac{\delta R}{x^2}}}.$$

$$\therefore \frac{1}{d_m^2} = \frac{\Sigma -n \cdot x^{n-3}}{\bar{x}^n} \cdot e^{-(x/\bar{x})^n} \delta x,$$

or 
$$\frac{1}{d_m^2} = \int_0^\infty \frac{-n x^{n-3}}{\bar{x}^n} \cdot e^{-(x/\bar{x})^n} dx. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Transforming this by the substitution  $y = \left(\frac{x}{\bar{x}}\right)^n$  gives

$$\frac{1}{d_m^2} = \frac{1}{\bar{x}^2} \int_0^\infty e^{-y} y^{-2/n} \cdot dy.$$

It may now be seen that the integral is a standard mathematical form and can be expressed in terms of gamma functions, whence

$$\frac{1}{d_m^2} = \frac{1}{\bar{x}^2} \Gamma\left(1 - \frac{2}{n}\right)$$

or 
$$d_m = \frac{x}{\sqrt{\Gamma\left(1 - \frac{2}{n}\right)}} = \frac{\bar{x}}{F(n)} \text{ say.} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In practice, it has been found that atomising nozzles give sprays having values of  $n$  lying between two and four. Fig. 1 shows a plot of  $F(n)$  for this range of values. The use of this graph reduces the computation of mean diameter to a very simple operation.

#### [6. MEAN DIAMETER OF A BURNING SPRAY.]

The mean diameter evaluated in Sec. 5 is for the injected spray and is related to the initial rate of evaporation of the spray, but it does not bear any obvious relation to the rate at which evaporation will be proceeding some time after evaporation has begun. For the subsequent history of the spray it is necessary to consider the behaviour of the individual drop.

In a combustion chamber where burning is steady, liquid fuel is constantly being injected while liquid fuel is also disappearing in the form of vapour, and the rates at which these two processes occur must be equal. It follows that there is present during steady combustion some constant volume of liquid evaporating at a rate equal to the rate of injection of the fuel. Clearly, the properties of this volume of fuel determine combustion characteristics, rather than those of the injected spray, except in so far as this spray is derived from the injected spray by evaporation and constant replacement. For convenience, the ratio of this volume to the volume injected per second will be known as the specific volume of the burning spray. The following sections are given to an examination of this specific volume.

## 6.1. EVAPORATION OF A DROP.

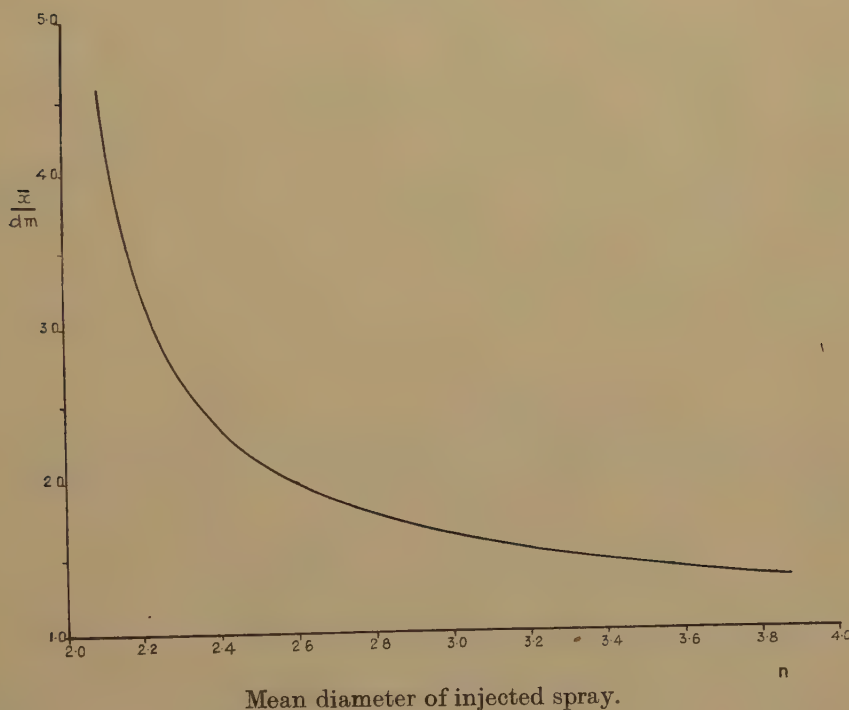
Since the rate of evaporation of a drop is proportional to its diameter, the law of evaporation for drops may be written

$$\frac{d}{dt}(x_t^3) = -Kx_t,$$

where  $x_t$  is the diameter at time  $t$ .

$K$  is a constant independent of drop size, dependent on humidity, etc., and assumed to be equal for all drops.

Fig. 1.



Then 
$$3x_t^2 \frac{d(x_t)}{dt} = -Kx_t \quad \text{or} \quad 3x_t d(x_t) = -K dt.$$

If  $x_0$  is the diameter at  $t=0$  integration gives

$$x_t^2 - x_0^2 = \frac{-2}{3} Kt.$$

For simplicity this is written as

$$x_t^2 = x_0^2 - \lambda t \quad \dots \dots \dots (8)$$

Henceforward,  $\lambda$  will be called the evaporation constant.

## 6.2. MEAN DIAMETER AND SPECIFIC VOLUME IN COMBUSTION.

Let the spray be injected at a volume rate  $V$  per second. As before, the small volume fraction of the injected spray in which the drops may be taken as of diameter  $x$  is

$$\delta R = -\frac{nx^{n-1}}{\bar{x}^n} e^{-(x/\bar{x})^n} \delta x.$$

Now the volume of one drop of diameter  $x$  is  $\frac{\pi}{6} x^3$ .

Therefore, the number of drops per unit volume injected, of size  $x$  is

$$\frac{\delta R}{\frac{\pi}{6} x^3} = \frac{6}{\pi} (-n) \frac{x^{n-4}}{\bar{x}^n} e^{-(x/\bar{x})^n} \delta x. \quad \dots \quad (9)$$

When a steady state of combustion has been established there will be drops of all sizes in the combustion chamber. If, at one instant, we consider all the drops of one size, we can account for them either as drops injected at that moment and of that size, or as the remains of larger drops injected earlier, which have evaporated down to that size. So at any instant consider the drops remaining from the spray injected  $t$  seconds earlier. The drops which are now of size  $x$  were then of size  $\sqrt{x^2 + \lambda t}$  (from 8).

Therefore, the number of drops now of size  $x$  per unit volume injected then was (from 9)

$$\frac{6}{\pi} (-n) \frac{(\sqrt{x^2 + \lambda t})^{n-4}}{\bar{x}^n} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x. \quad \dots \quad (10)$$

But the number of drops of a given initial size remains constant as they evaporate. Therefore, the total number of drops of size  $x$  in the steady state is

$$\sum_{t=0}^{\infty} \frac{6}{\pi} (-n) \frac{(\sqrt{x^2 + \lambda t})^{n-4}}{\bar{x}^n} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x \cdot V \cdot \delta t. \quad \dots \quad (11)$$

The total volume of drops of size  $x$  is

$$x^3 \sum_{t=0}^{\infty} (-n) \frac{(\sqrt{x^2 + \lambda t})^{n-4}}{\bar{x}^n} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x \cdot V \cdot \delta t. \quad \dots \quad (12)$$

The total volume of liquid present during combustion is

$$\sum_{x=0}^{\infty} \sum_{t=0}^{\infty} (-n) \frac{x^3}{\bar{x}^n} (\sqrt{x^2 + \lambda t})^{n-4} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x \cdot V \cdot \delta t. \quad \dots \quad (13)$$

Reference to equation (4) shows that the mean diameter of the specific volume is

$$\sqrt{\frac{\sum_{x=0}^{\infty} \sum_{t=0}^{\infty} (-n) \frac{x^3}{\bar{x}^n} (\sqrt{x^2 + \lambda t})^{n-4} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x \cdot V \cdot \delta t}{\sum_{x=0}^{\infty} \sum_{t=0}^{\infty} (-n) \frac{x}{\bar{x}^n} (\sqrt{x^2 + \lambda t})^{n-4} e^{-(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}})^n} \delta x \cdot V \cdot \delta t}}. \quad \dots \quad (14)$$



Expressions (13) and (14) are evaluated in Appendix I., whence specific volume

$$= \frac{\bar{x}^2}{2\lambda} \Gamma\left(1 + \frac{2}{n}\right) \quad \dots \dots \dots (15)$$

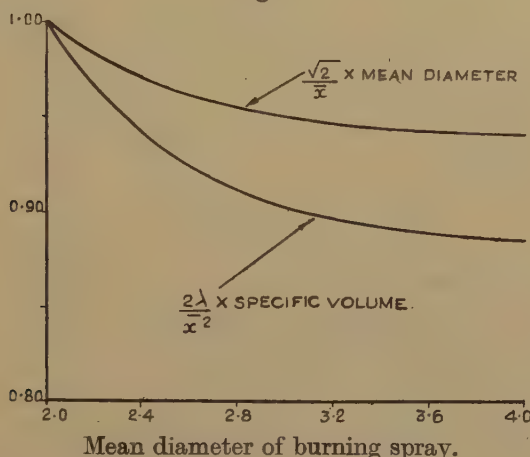
and mean diameter

$$= \bar{x} \sqrt{\frac{\Gamma\left(1 + \frac{2}{n}\right)}{2}} \quad \dots \dots \dots (16)$$

A graph of  $\Gamma\left(1 + \frac{2}{n}\right)$  and  $\sqrt{\Gamma\left(1 + \frac{2}{n}\right)}$  is given in fig. 2. It is apparent that the distribution constant does not have a great effect on these quantities. For practical purposes, it would be sufficient to write

$$\text{Mean diameter} = \frac{\bar{x}}{\sqrt{2}}.$$

Fig. 2.



## 7. COMBUSTION INTENSITY.

It is of practical interest to know how spray characteristics will affect combustion intensity, or alternatively, how they affect the size of combustion chamber when a given quantity of fuel is to be burnt. The definition of intensity of combustion would be quantity of fuel evaporated per unit space per unit time. If it is assumed that all the spray is evaporated, intensity then becomes proportional to the quantity evaporated divided by the time necessary to evaporate the largest drop. Unfortunately, the distribution law does not bar the presence of a very small volume of liquid composed of very large drops ; while this does not question the validity of the law, it does make a purely mathematical approach to intensity impractical.

A more realistic method is to relate combustion intensity to the specific

volume in combustion. It has been shown that in order to evaporate a quantity of liquid  $V$  per second, a volume

$$\frac{V\bar{x}^2}{2\lambda} \Gamma\left(1 + \frac{2}{n}\right)$$

must be constantly evaporating. The problem in combustion is to maintain a proportionate mixture of fuel and air throughout, and we may infer that during evaporation a volume of air proportional to the volume of liquid must be present, and that the volume of evaporation space is therefore proportional to the volume of liquid.

It seems reasonable then, that in the first analysis flame intensity is inversely proportional to specific volume, or intensity

$$\propto \frac{2\lambda}{\bar{x}^2 \Gamma\left(1 + \frac{2}{n}\right)}.$$

### 8. INCOMPLETE EVAPORATION OF A SPRAY.

The last quantity to be discussed, and perhaps the most important one, is the fraction of an injected spray remaining after a given time. It would be expected that combustion losses would be most sensitive to this, since in a combustion chamber there is a definite maximum time allowed.

As before, let a volume  $V$  be injected per second. From (9) the number of drops of size  $x$  injected per second is

$$\frac{6}{\pi}(-n) \frac{x^{n-4}}{\bar{x}^n} e^{-(x/\bar{x})^n} \delta x \cdot V.$$

After  $t$  seconds, these drops have decreased in diameter from  $x$  to  $\sqrt{x^2 - \lambda t}$  and the volume of each drop is then  $\frac{\pi}{6}(x^2 - \lambda t)^{3/2}$ .

Thus, the volume of drops initially of size  $x$  is now

$$(-n) \frac{x^{n-4}}{\bar{x}^n} \cdot e^{-(x/\bar{x})^n} (x^2 - \lambda t)^{3/2} \delta x \cdot V. \quad \dots \quad (17)$$

If the time allowed for evaporation is  $T$  seconds, the drops initially of size  $\sqrt{\lambda T}$  are just disappearing and only drops initially greater than  $\sqrt{\lambda T}$  are now contributing to the remaining volume. Therefore, the total volume remaining after  $T$  seconds is

$$\sum_{x=\sqrt{\lambda T}}^{x=\infty} (-n) \frac{x^{n-4}}{\bar{x}^n} (x^2 - \lambda T)^{3/2} e^{-(x/\bar{x})^n} \delta x \cdot V, \quad \dots \quad (18)$$

The volume fraction remaining is

$$\sum_{x=\sqrt{\lambda T}}^{x=\infty} (-n) \frac{x^{n-4}}{\bar{x}^n} (x^2 - \lambda T)^{3/2} e^{-(x/\bar{x})^n} \delta x.$$

or

$$\int_{\sqrt{\lambda T}}^{\infty} (-n) \frac{x^{n-4}}{\bar{x}^n} (x^2 - \lambda T)^{3/2} e^{-(x/\bar{x})^n} dx. \quad \dots \quad (19)$$

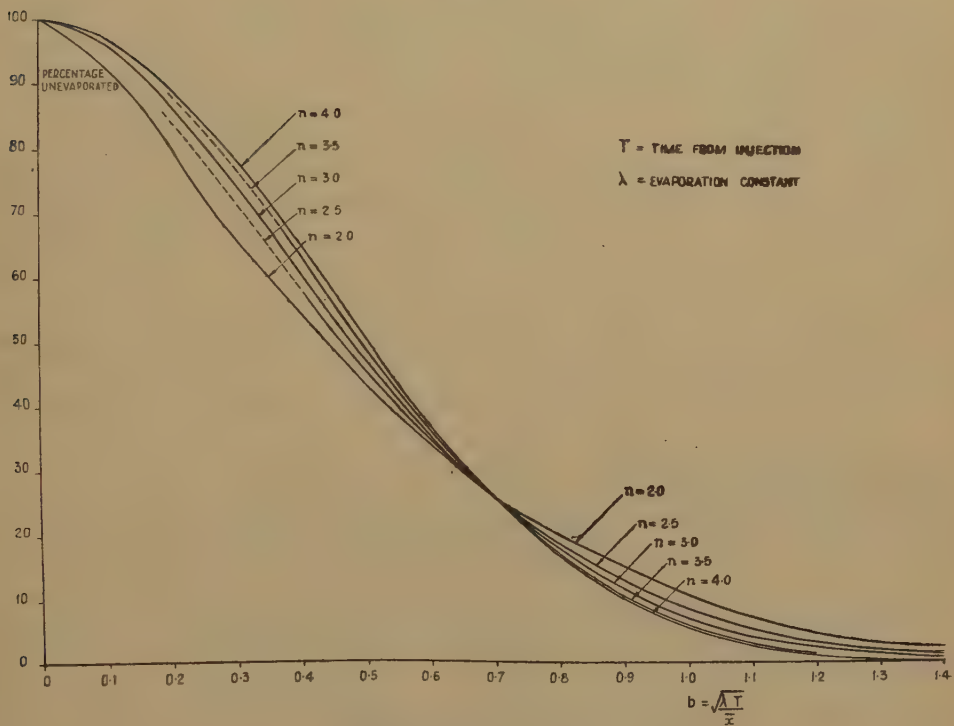
Attempts to integrate this expression (19) have not been successful, neither could a suitable expansion be developed.

However, by substituting  $x/\bar{x} = b \sec \theta$ , where  $b = \frac{\sqrt{\lambda T}}{\bar{x}}$ , expression (19) may be transformed to

$$3 \int_0^{\pi/2} e^{-(b \sec \theta)^n} \sin^2 \theta \cos \theta d\theta. \quad (20)$$

This expression is more suitable for graphical integration, which has been carried out for a suitable range of values of  $b$  and  $n$ . The results are

Fig. 3.



Evaporation of a spray.

shown in fig. 3, where the unevaporated residue (per cent.) is plotted against  $\frac{\sqrt{\lambda T}}{\bar{x}}$  for values of  $n$  between two and four.

These curves are most instructive in the following respect, they show clearly how misleading it is to rate fuel sprays on their initial rates of evaporation when injected. For although the times for evaporation of a given quantity are much more sensitive to changes in  $\bar{x}$  than in  $n$ , yet the distribution constant  $n$  does have a great effect on the time necessary to evaporate the last few per cent. of the spray. A small value of  $n$  gives a small mean diameter of the injected spray (ref. fig. 2) and correspondingly

a high initial rate of evaporation, but after about 75 per cent. of the liquid has evaporated the effect of  $n$  is inverted and the rate of evaporation is slower for small values of  $n$ . It is now seen that  $\bar{x}$  should be as small as possible, and  $n$  large, for quick evaporation of the whole spray, even though a large value of  $n$  means a slower initial rate of evaporation.

#### CONCLUSIONS.

The evaporation of a liquid spray has been examined and the effects of droplet size and distribution made clear. The examination has shown that by far the most important of the spray parameters is  $\bar{x}$  the size constant, while  $n$ , the distribution constant, has its chief effect on the time required for the evaporation of the last one or two per cent. It is desirable that a liquid fuel spray should have a small value of  $\bar{x}$  and a large value of  $n$ , even though this condition for  $n$  gives a slower initial rate of evaporation of the injected spray. Simply, it is not enough merely to consider the injected spray alone. The graphs appended allow simple calculation of the rates of evaporation of liquid sprays and show how the liquid volume of a spray varies with time during its life. These graphs should be of use in correlating data on the flame lengths from burners and on combustion losses.

#### ACKNOWLEDGMENT.

The author wishes to thank Mrs. E. M. Cohen for her help with the integration contained in the Appendix.

#### REFERENCES.

- (1) 'Encyclopædia Britannica,' "Diffusion."
- (2) Namekawa and Takahashi, "Note on the Evaporation of Small Water Drops," Physico-Mathematical Society of Japan, 1932.
- (3) Takahashi, 'Experiments on the Evaporation of Water Drops.'

#### APPENDIX I.

$$\begin{aligned} \text{Specific volume} &= \sum_{x=0}^{x=\infty} \sum_{t=0}^{t=\infty} (-n) \frac{x^3}{x^n} (\sqrt{x^2 + \lambda t})^{n-4} e^{-\left(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}}\right)^n} \delta x \delta t \\ &= \int_{x=0}^{\infty} \int_{t=0}^{\infty} (-n) \frac{x^3}{x^n} (\sqrt{x^2 + \lambda t})^{n-4} e^{-\left(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}}\right)^n} dx dt. \end{aligned}$$

$$\begin{aligned} \text{Substitute} \quad x &= \bar{x} R \cos \theta, \\ \lambda t &= \bar{x}^2 R^2 \sin^2 \theta. \end{aligned}$$

$$\begin{aligned} \text{Then} \quad \frac{\partial x}{\partial R} &= \bar{x} \cos \theta, \\ \frac{\partial x}{\partial \theta} &= -\bar{x} R \sin \theta, \\ \lambda \frac{\partial t}{\partial R} &= 2\bar{x}^2 R \sin^2 \theta, \\ \lambda \frac{\partial t}{\partial \theta} &= 2\bar{x}^2 R^2 \sin \theta \cos \theta. \end{aligned}$$



Then

$$\begin{aligned}
 dx \, dt &= J \left| \frac{x \, t}{R \, \theta} \right| dR \, d\theta, \\
 &= \left( \frac{\partial x}{\partial R} \cdot \frac{\partial t}{\partial \theta} - \frac{\partial x}{\partial \theta} \frac{\partial t}{\partial R} \right) dR \, d\theta, \\
 &= \frac{2x^3 R^2}{\lambda} \sin \theta \cos \theta.
 \end{aligned}$$

So specific volume

$$\begin{aligned}
 &= \int_0^\infty \int_0^{\pi/2} \frac{(-n) \bar{x}^3 R^3 \cos^3 \theta}{\bar{x}^n} \bar{x}^{n-4} R^{n-4} e^{-R^n} \cdot \frac{2 \bar{x}^3 R^2}{\lambda} \sin \theta \, dR \, d\theta, \\
 &= \frac{-2n \bar{x}^2}{\lambda} \int_0^\infty \int_0^{\pi/2} R^{n+1} e^{-R^n} \cos^3 \theta \sin \theta \, dR \, d\theta, \\
 &= \frac{-2n \bar{x}^2}{\lambda} \int_0^\infty e^{-R^n} R^{n+1} dR \int_0^{\pi/2} \cos^3 \theta \sin \theta \, d\theta, \\
 &= \frac{-2n \bar{x}^2}{2\lambda} \int_0^\infty e^{-R^n} R^{n+1} dR.
 \end{aligned}$$

Substituting  $R^n = y$ , the integral reduces to

$$\begin{aligned}
 &\frac{-\bar{x}^2}{2\lambda} \int_0^\infty e^{-y} y^{2/n} dy, \\
 &= \frac{-\bar{x}^2}{2\lambda} \Gamma\left(1 + \frac{2}{n}\right).
 \end{aligned}$$

Similarly the denominator in (14) is

$$\begin{aligned}
 &\sum_{x=0}^{x=\infty} \sum_{t=0}^{t=\infty} (-n) \frac{x}{\bar{x}} \sqrt{(x^2 + \lambda t)^{n-4}} e^{-\left(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}}\right)^n} \delta x \, \delta t, \\
 &\int_{x=0}^{x=\infty} \int_{t=0}^{t=\infty} (-n) \frac{x}{\bar{x}^n} \sqrt{(x^2 + \lambda t)^{n-4}} e^{-\left(\frac{\sqrt{x^2 + \lambda t}}{\bar{x}}\right)^n} dx \, dt.
 \end{aligned}$$

Making the same substitution as before, this integral reduces to

$$\begin{aligned}
 &\frac{-2n}{\lambda} \int_0^\infty \int_0^{\pi/2} e^{-R^n} R^{n-1} \sin \theta \cos \theta \, dR \, d\theta, \\
 &= \frac{-2n}{\lambda} \int_0^\infty -R^n R^{n-1} dR \int_0^{\pi/2} \sin \theta \cos \theta \, d\theta, \\
 &= \frac{-1}{\lambda} \int_0^\infty d(e^{-R^n}), \\
 &= \frac{-1}{\lambda}.
 \end{aligned}$$

Thus specific volume  $= \frac{\bar{x}^2}{2\lambda} \Gamma(1 + 2/n).$

Mean diameter  $= \frac{\sqrt{\bar{x}^2 \Gamma(1 + 2/n)}}{2}.$

*XVI. An Escalator Process for the Solution of Linear  
Simultaneous Equations.*

By J. MORRIS, B.A.\*

[Received November 12, 1945.]

SUMMARY.

The method given in this paper for the solution of linear simultaneous equations is on the lines of the escalator process originally devised for the solution of Lagrangian frequency equations.

The process does not involve the usual devices of elimination or of the evaluation of determinants, neither does it involve iteration or the like.

The escalator process is based essentially on an artifice for the successive introduction of each of the variables involved by definite self-contained stages.

An important feature of the process is that powerful checks are available at each stage to ensure accuracy and if necessary to adjust it.

---

1. INTRODUCTION.

It is a matter of common experience that the various processes for the solution of linear simultaneous equations which arise in a variety of practical problems are extremely sensitive in operations; and that unless an adequate number of significant figures is taken into account in the course of the computation, considerable errors may accrue in the results. It seems that there is a fallacy in the assumption that it is superfluous to work to more significant figures than occur in the coefficients or constants in the equations. Let us take for example the following set of four equations devised by Mr. T. S. Wilson, viz. :—

$$\begin{aligned} 5x + 7y + 6z + 5u &= 23, \\ 7x + 10y + 8z + 7u &= 32, \\ 6x + 8y + 10z + 9u &= 33, \\ 5x + 7y + 9z + 10u &= 31. \end{aligned}$$

The exact solutions are  $x=y=z=u=1$ .

---

\* Communicated by the Director, Royal Aircraft Establishment, Farnborough.

Nevertheless, it may be verified that

$x.$	$y.$	$z.$	$u.$	$d_1.$	$d_2.$	$d_3.$	$d_4.$
14.6	-7.2	-2.5	3.1	23.1	31.9	32.9	31.1
2.36	0.18	0.65	1.21	23.01	31.99	32.99	31.01
1.136	0.918	0.965	1.021	23.001	31.999	32.999	31.001

where the  $d$ 's are the results of the substitution in the left-hand sides of the equations of the above sets of values of the variables, and would be the appropriate constants with the given coefficients for the respective tabulated values of the "unknowns."

We can readily see the reason for this sensitivity if we derive the solutions of the given equations in terms of the  $d$ 's as follows:—

$$\begin{aligned}x &= 68d_1 - 41d_2 - 17d_3 + 10d_4, \\y &= -41d_1 + 25d_2 + 10d_3 - 6d_4, \\z &= -17d_1 + 10d_2 + 5d_3 - 3d_4, \\u &= 10d_1 - 6d_2 - 3d_3 + 2d_4;\end{aligned}$$

from which it is apparent that small changes in the constants may result in relatively large variations in the values of the variables.

Reciprocally, if we regard the  $d$ 's as the variables and  $x, y, z, u$  as the constants in the transformed equations, then it is clear that small variations in the variables may result in large variations in the constants. Thus the transformed equations are also sensitive.

Equations of this kind are called "ill-conditioned."

To illustrate how ill-conditioned equations may crop up, let us consider the quadric

$$f(x, y) = \frac{1}{2}(a_{11}x^2 + 2a_{12}xy + a_{22}y^2 - 2d_1x - 2d_2y + c) = 0, \quad \dots (1)$$

which, if  $a_{11}, a_{22}, (a_{11}a_{22} - a_{12}^2)$  are positive, will represent an ellipse.

If now we wish to find the centre of the quadric, we form two linear simultaneous equations for its co-ordinates as follows:—

$$\partial f / \partial x = a_{11}x + a_{12}y - d_1 = 0, \quad \dots (2)$$

$$\partial f / \partial y = a_{12}x + a_{22}y - d_2 = 0. \quad \dots (3)$$

These equations will represent the diameters of the quadric which connect pairs of points of contact of tangents parallel respectively to the  $x$ - and  $y$ -axes.

If  $a_{11}, a_{12}, a_{22}$ , are such that these diametral lines cut at a very acute angle, then the equations for them will be ill-conditioned; and this will occur if  $(a_{11}a_{22} - a_{12}^2)$  is small compared with  $a_{11}$  or  $a_{22}$ .

Now in practical problems we obtain in a precisely similar manner equations like (2) and (3) from a quadratic function like (1), the resulting equations being conditioned by the relative values of the

quantities corresponding to the coefficients  $a_{11}$ ,  $a_{12}$ ,  $a_{22}$ , and thus the equations to be solved may in certain cases become ill-conditioned.

If the linear simultaneous equations have integral coefficients, and if we can avoid any form of approximation, then the solution will be accurate however the equations are conditioned. Ill-conditioning manifests itself where recourse has to be made to approximation, in the course of which the number of significant figures taken into account is inadequate. Ill-conditioning becomes very apparent in iteration processes, no matter to how many significant figures we work, since the convergence may be so slow as to render the process abortive. For instance, the usual iteration process<sup>(1)</sup>, when applied to Mr. Wilson's set of equations, follows the scheme

$$5x_{n+1} - 7y_n + 6z_n + 5u_n = 23,$$

$$7x_{n+1} + 10y_{n+1} + 8z_n + 7u_n = 32,$$

$$6x_{n+1} + 8y_{n+1} + 10z_{n+1} + 9u_n = 33,$$

$$5x_{n+1} + 7y_{n+1} + 9z_{n+1} + 10u_{n+1} = 31.$$

It may be shown that the convergence of the process depends on the values of the  $\lambda$  roots of the appropriate characteristic equation, which in the particular case is

$$\begin{vmatrix} 5\lambda & 7 & 6 & 5 \\ 7\lambda & 10\lambda & 8 & 7 \\ 6\lambda & 8\lambda & 10\lambda & 9 \\ 5 & 7 & 9 & -10 \end{vmatrix} = 0.$$

The smaller the values of the  $\lambda$ 's the more rapid will be the convergence; if any of the  $\lambda$ 's exceed unity (or if imaginary their moduli exceed unity), then the iteration process will be divergent.

For the equations in question the values of the  $\lambda$ 's are found to be to three significant figures

$$\lambda_1 = 0.997, \lambda_2 = 0.837, \lambda_3 = 0.604.$$

Since the convergency of the iteration process referred to depends on the differences of successive values of the variables iterated, *e.g.*  $x_{n+1} - x_n$ , whose magnitudes involve geometrical powers of the  $\lambda$ 's, it is apparent that the process in the particular case will be slow and sluggish.

In this paper a novel method is put forward for the solution of linear simultaneous equations. The method follows closely on the lines of the escalator<sup>(2)</sup> process, which was originally devised for the solution of Lagrangian frequency equations. The escalator process is based essentially on an artifice for the successive introduction of each of the variables involved by definite self-contained stages, in which the solutions of a set of equations of say the fourth order is obtained *inter alia* in terms of the solutions for the preceding third order equations, and so on.



The escalator process does not involve the usual devices of elimination or the evaluation of determinants ; neither does it involve iteration or such like approximation. An important feature of the process is that powerful checks are available at each stage to ensure accuracy and, if necessary, to adjust it.

Although in a general sense the escalator process is applicable to equations of any form, it would appear to be advantageous in cases where the equations are symmetrical, and particularly in which the coefficients are coefficients of a positive homogeneous quadratic function. Equations which arise from practical problems are usually of this form ; if not they can readily be converted into such by "normalization." By normalization we mean the following procedure. Consider, for example, the three equations

$$a_1x+b_1y+c_1z=d_1,$$

$$a_2x+b_2y+c_2z=d_2,$$

$$a_3x+b_3y+c_3z=d_3.$$

Multiply them in turn by  $a_1, a_2, a_3$ , and add ; then in turn by  $b_1, b_2, b_3$ , and add ; and finally in turn by  $c_1, c_2, c_3$ , and add. The resulting equations will have coefficients pertaining to the following positive homogeneous quadratic function, viz. :—

$$f(xyz)=(a_1x+b_1y+c_1z)^2+(a_2x+b_2y+c_2z)^2+(a_3x+b_3y+c_3z)^2.$$

Finally, it would appear that the escalator process is not unduly sensitive, even when the equations to be solved are ill-conditioned as regards the usual processes for solution.

When, however, the equations are very well conditioned, then it would appear that the iteration process has decided advantages.

#### ACKNOWLEDGMENT.

The Author wishes to acknowledge the invaluable assistance he has received from Mr. T. S. Wilson and Miss E. Giles, both in formulation of the analysis and in the organization of the means for its numerical application.

### 2. THEORY OF THE PROCESS.

#### 2.1. *Symmetrical Equations.*

Let us consider the following four symmetrical linear simultaneous equations, viz. :—

$$a_{11}x+a_{12}y+a_{13}z+a_{14}u=d_1, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$a_{12}x+a_{22}y+a_{23}z+a_{24}u=d_2, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$a_{13}x+a_{23}y+a_{33}z+a_{34}u=d_3, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$a_{14}x+a_{24}y+a_{34}z+a_{44}u=d_4. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Let  $a_{11}x_2+a_{12}=0, \quad a_{12}x_2+a_{22}=q_2 ; \quad . \quad . \quad . \quad . \quad . \quad (5)$

then

$$x_2 = -a_{12}/q_1, \text{ where } q_1 = a_{11}. \quad (6)$$

Next let

$$a_{11}x_3 + a_{12}y_3 + a_{13} = 0, \quad (7)$$

$$a_{12}x_3 + a_{22}y_3 + a_{23} = 0, \quad (8)$$

$$a_{13}x_3 + a_{23}y_3 + a_{33} = q_3. \quad (9)$$

Using the previous results, we find

$$y_3 = -(a_{13}x_2 + a_{23})/q_2, \quad (10)$$

$$x_3 = -[a_{13}/q_1 + (a_{13}x_2 + a_{23})x_2/q_2], \quad (11)$$

$$q_3 = a_{33} - a_{13}^2/q_1 - (a_{13}x_2 + a_{23})^2/q_2. \quad (12)$$

Next let

$$a_{11}x_4 + a_{12}y_4 + a_{13}z_4 + a_{14} = 0, \quad (13)$$

$$a_{12}x_4 + a_{22}y_4 + a_{23}z_4 + a_{24} = 0, \quad (14)$$

$$a_{13}x_4 + a_{23}y_4 + a_{33}z_4 + a_{34} = 0, \quad (15)$$

$$a_{14}x_4 + a_{24}y_4 + a_{34}z_4 + a_{44} = q_4; \quad (16)$$

then it may be shown that

$$z_4 = -(a_{14}x_3 + a_{24}y_3 + a_{34})/q_3, \quad (17)$$

$$y_4 = -[(a_{14}x_2 + a_{24})/q_2 + (a_{14}x_3 + a_{24}y_3 + a_{34})y_3/q_3], \quad (18)$$

$$x_4 = -[a_{14}/q_1 + (a_{14}x_2 + a_{24})x_2/q_2 + (a_{14}x_3 + a_{24}y_3 + a_{34})x_3/q_3], \quad (19)$$

$$q_4 = a_{44} - a_{14}^2/q_1 - (a_{14}x_2 + a_{24})^2/q_2 - (a_{14}x_3 + a_{24}y_3 + a_{34})^2/q_3; \quad (20)$$

or if we write

$$a_{14}x_3 + a_{24}y_3 + a_{34} = (a_{14})_3,$$

$$a_{14}x_2 + a_{24} = (a_{14})_2, \quad (a_{14})_1 = a_{14}; \quad (21)$$

then

$$z_4 = -(a_{14})_3/q_3, \quad (22)$$

$$y_4 = -[(a_{14})_2/q_2 + (a_{14})_3y_3/q_3], \quad (23)$$

$$x_4 = -[(a_{14})_1/q_1 + (a_{14})_2x_2/q_2 + (a_{14})_3x_3/q_3], \quad (24)$$

$$q_4 = a_{44} - (a_{14})_1^2/q_1 - (a_{14})_2^2/q_2 - (a_{14})_3^2/q_3. \quad (25)$$

If now we designate the solution of the equations (1), (2), (3), (4), by  $x_4, \bar{y}_4, \bar{z}_4, \bar{u}_4$ , it follows that

$$\bar{u}_4 = (\bar{d}_1)_4/q_4, \quad (26)$$

$$\bar{z}_4 = (\bar{d}_1)_3/q_3 + (\bar{d}_1)_4z_4/q_4, \quad (27)$$

$$\bar{y}_4 = (\bar{d}_1)_2/q_2 + (\bar{d}_1)_3y_3/q_3 + (\bar{d}_1)_4y_4/q_4, \quad (28)$$

$$\bar{x}_4 = (\bar{d}_1)_1/q_1 + (\bar{d}_1)_2x_2/q_2 + (\bar{d}_1)_3x_3/q_3 + (\bar{d}_1)_4x_4/q_4, \quad (29)$$

where

$$(\bar{d}_1)_1 = \bar{d}_1, \quad (\bar{d}_1)_2 = \bar{d}_1x_2 + \bar{d}_2, \quad (30)$$

$$(\bar{d}_1)_3 = \bar{d}_1x_3 + \bar{d}_2y_3 + \bar{d}_3, \quad (31)$$

$$p = \bar{d}_{14}x_4 + \bar{d}_2y_4 + \bar{d}_3z_4 + \bar{d}_4. \quad (32)$$

We notice that with the notation adopted,

$$q_1=(a_{11})_1, \quad q_2=(a_{12})_2, \quad q_3=(a_{13})_3, \quad q_4=(a_{14})_4,$$

and so on.

## 2.2. *Unsymmetrical Equations.*

If the equations are unsymmetrical as follows :—

$$a_{11}x+a_{12}y+a_{13}z+a_{14}u=d_1, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$a_{21}x+a_{22}y+a_{23}z+a_{24}u=d_2, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$a_{31}x+a_{32}y+a_{33}z+a_{34}u=d_3, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$a_{41}x+a_{42}y+a_{43}z+a_{44}u=d_4; \quad . \quad . \quad . \quad . \quad . \quad (4)$$

we consider simultaneously the transposed equations

$$a_{11}x'+a_{21}y'+a_{31}z'+a_{41}u'=d_1, \quad . \quad . \quad . \quad . \quad . \quad (1)'$$

$$a_{12}x'+a_{22}y'+a_{32}z'+a_{42}u'=d_2, \quad . \quad . \quad . \quad . \quad . \quad (2)'$$

$$a_{13}x'+a_{23}y'+a_{33}z'+a_{43}u'=d_3, \quad . \quad . \quad . \quad . \quad . \quad (3)'$$

$$a_{14}x'+a_{24}y'+a_{34}z'+a_{44}u'=d_4, \quad . \quad . \quad . \quad . \quad . \quad (4)'$$

Thus corresponding to, say,

$$a_{11}x_3+a_{12}y_3+a_{13}=0, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$a_{21}x_3+a_{22}y_3+a_{23}=0, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$a_{31}x_3+a_{32}y_3+a_{33}=q_3, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

we shall have

$$a_{11}x_3'+a_{21}y_3'+a_{31}=0, \quad . \quad . \quad . \quad . \quad . \quad (5)'$$

$$a_{12}x_3'+a_{22}y_3'+a_{32}=0, \quad . \quad . \quad . \quad . \quad . \quad (6)'$$

$$a_{13}x_3'+a_{32}y_3'+a_{33}=q_3'. \quad . \quad . \quad . \quad . \quad . \quad (7)'$$

Clearly  $q_3'=q_3$ .

Also, we find

$$x_3=-[a_{13}/q_1+(a_{13}x_2'+a_{23})x_2/q_2], \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$y_3=-(a_{13}x_2'+a_{23})/q_2; \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$x_3'=-[a_{31}/q_1+(a_{31}x_2+a_{32})x_2'/q_2], \quad . \quad . \quad . \quad . \quad . \quad (8)'$$

$$y_3'=-[a_{31}x_2+a_{32})/q_2; \quad . \quad . \quad . \quad . \quad . \quad (9)'$$

$$q_3=a_{33}-a_{13}a_{31}/q_1-(a_{13}x_2'+a_{23})(a_{31}x_2+a_{32})/q_2. \quad . \quad . \quad . \quad (10)$$

The solution to the unsymmetrical equations (1), (2), (3), (4), will be found to be,

$$\bar{u}_4=(d_1')_4/q_4, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$\bar{z}_4=(d_1')_3/q_3+(d_1')_4z_4/q_4, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$\bar{y}_4=(d_1')_2/q_2+(d_1')_3y_3/q_3+(d_1')_4y_4/q_4, \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$\bar{x}_4=(d_1')_1/q_1+(d_1')_2x_2/q_2+(d_1')_3x_3/q_3+(d_1')_4x_4/q_4; \quad . \quad . \quad (14)$$





where

$$(d_1')_4 = d_1x_4' + d_2y_4' + d_3z_4' + d_4,$$

$$(d_1)_4 = d_1x_4 + d_2y_4 + d_3z_4 + d_4.$$

Although we have dealt only with equations up to the fourth order, is clear that the process is quite general and it is evident how the expressions for the appropriate quantities for succeeding orders proceed

### 2.3. *Changing a Row and Column.*

Suppose we have the solutions by the escalator process of the fourth order equations (1), (2), (3), (4), of para. 1, in which circumstances we shall have available the values of the  $\alpha_{rs}$ 's in the inverse equations, viz. :—

$$x = \alpha_{11}d_1 + \alpha_{12}d_2 + \alpha_{13}d_3 + \alpha_{14}d_4, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$y = \alpha_{12}d_1 + \alpha_{22}d_2 + \alpha_{23}d_3 + \alpha_{24}d_4, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$z = \alpha_{13}d_1 + \alpha_{23}d_2 + \alpha_{33}d_3 + \alpha_{34}d_4, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$u = \alpha_{14}d_1 + \alpha_{24}d_2 + \alpha_{34}d_3 + \alpha_{44}d_4, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Suppose next that we wish to know what the effect on the solutions will be if we change the elements of the second row and column, i. e. if we change  $a_{12}, a_{22}, a_{23}, a_{24}, d_2$  to  $a_{12}', a_{22}', a_{23}', a_{24}', d_2'$ .

We first find what value of  $d_2$  will make  $y=0$  in (2), the other values of the  $d$ 's being unchanged. We then substitute this value of  $d_2$  in (1), (3), (4), and thus obtain the following :—

$$x = \beta_{11}d_1 + \beta_{13}d_3 + \beta_{14}d_4, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$z = \beta_{13}d_1 + \beta_{33}d_3 + \beta_{34}d_4, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$u = \beta_{14}d_1 + \beta_{34}d_3 + \beta_{44}d_4, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where

$$\beta_{11} = (\alpha_{11} - \alpha_{12}^2/\alpha_{22}), \quad \beta_{13} = (\alpha_{13} - \alpha_{12}\alpha_{23}/\alpha_{22}),$$

$$\beta_{14} = (\alpha_{14} - \alpha_{12}\alpha_{24}/\alpha_{22}), \quad \beta_{33} = (\alpha_{33} - \alpha_{23}^2/\alpha_{22}),$$

$$\beta_{34} = (\alpha_{34} - \alpha_{23}\alpha_{24}/\alpha_{22}), \quad \beta_{44} = (\alpha_{44} - \alpha_{24}^2/\alpha_{22}).$$

These values of  $x, z, u$ , will then be the solution of (1), (3), (4), with  $y=0$ .

Next write the new equations in the form

$$a_{11}x + a_{12}'y + a_{13}z + a_{14}u = d_1, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$a_{12}'x + a_{22}'y + a_{23}'z + a_{24}'u = d_2', \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$a_{13}x + a_{23}'y + a_{33}z + a_{34}u = d_3, \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$a_{14}x + a_{24}'y + a_{34}z + a_{44}u = d_4, \quad . \quad . \quad . \quad . \quad . \quad (11)$$

We have at once by (5), (6), (7), the solutions to the equations

$$a_{11}x + a_{13}z + a_{14}u = -a_{12}'y, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$a_{13}x + a_{33}z + a_{34}u = -a_{23}'y, \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$a_{14}x + a_{34}z + a_{44}u = -a_{24}'y, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$$\text{in the form} \quad x/y = -[\beta_{11}a_{12}' + \beta_{13}a_{23}' + \beta_{14}a_{24}'], \quad . \quad . \quad . \quad . \quad . \quad (15)$$

$$z/y = -[\beta_{13}a_{12}' + \beta_{33}a_{23}' + \beta_{34}a_{24}'], \quad . \quad . \quad . \quad . \quad . \quad (16)$$

$$u/y = -[\beta_{14}a_{12}' + \beta_{34}a_{23}' + \beta_{44}a_{24}'], \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Let us call these ratios  $r_{12}'$ ,  $r_{32}'$ ,  $r_{42}'$ .

Then from (8)  $r_{12}' + (9) + (10)r_{32}' + (11)r_{42}'$ , and having regard to (12), (13), (14), we obtain

$$\begin{aligned} (a_{12}'r_{12}' + a_{22}' + a_{23}'r_{32}' + a_{24}'r_{42}')\bar{y}_4' \\ = (d_1r_{12}' + d_2' + d_3r_{32}' + d_4r_{42}'). \quad . \quad . \quad (18) \end{aligned}$$

This gives the value of  $\bar{y}_4'$  required.

To find the values of  $\bar{x}_4'$ ,  $\bar{z}_4'$ ,  $\bar{u}_4'$ , we substitute the value of  $\bar{y}_4'$  as given by (18) in equations (8), (10), (11), and so obtain equations like (12), (13), (14), with the right-hand sides respectively.

$$d_1 - a_{12}'\bar{y}_4', \quad d_3 - a_{23}'\bar{y}_4', \quad d_4 - a_{24}'\bar{y}_4';$$

The solutions required for  $\bar{x}_4'$ ,  $\bar{z}_4'$ ,  $\bar{u}_4'$ , will then be given by (5), (6), (7), with the appropriate values of the  $d$ 's.

#### 2.4. Solution by Partition.

Suppose we have the six equations

$$a_{11}x + a_{12}y + a_{13}z + a_{14}u + a_{15}v + a_{16}w = d_1, \quad . \quad . \quad . \quad . \quad (1)$$

$$a_{12}x + a_{22}y + a_{23}z + a_{24}u + a_{25}v + a_{26}w = d_2, \quad . \quad . \quad . \quad . \quad (2)$$

$$a_{13}x + a_{23}y + a_{33}z + a_{34}u + a_{35}v + a_{36}w = d_3, \quad . \quad . \quad . \quad . \quad (3)$$

$$a_{14}x + a_{24}y + a_{34}z + a_{44}u + a_{45}v + a_{46}w = d_4, \quad . \quad . \quad . \quad . \quad (4)$$

$$a_{15}x + a_{25}y + a_{35}z + a_{45}u + a_{55}v + a_{56}w = d_5, \quad . \quad . \quad . \quad . \quad (5)$$

$$a_{16}x + a_{26}y + a_{36}z + a_{46}u + a_{56}v + a_{66}w = d_6; \quad . \quad . \quad . \quad . \quad (6)$$

we may proceed as follows:—

Writing equations (1), (2), (3), in the form

$$a_{11}x + a_{12}y + a_{13}z = \delta_1, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$a_{12}x + a_{22}y + a_{23}z = \delta_2, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$a_{13}x + a_{23}y + a_{33}z = \delta_3, \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where

$$\delta_1 = d_1 - (a_{14}u + a_{15}v + a_{16}w), \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\delta_2 = d_2 - (a_{24}u + a_{25}v + a_{26}w), \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$\delta_3 = d_3 - (a_{34}u + a_{35}v + a_{36}w); \quad . \quad . \quad . \quad . \quad . \quad (12)$$

we solve these by the escalator process in the form

$$x = \alpha_{11}\delta_1 + \alpha_{12}\delta_2 + \alpha_{13}\delta_3, \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$y = \alpha_{12}\delta_1 + \alpha_{22}\delta_2 + \alpha_{23}\delta_3, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$$z = \alpha_{13}\delta_1 + \alpha_{23}\delta_2 + \alpha_{33}\delta_3. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

We then substitute these values in (4), (5), (6), and thus form a set of third order equations, which we shall denote as

$$a_{11}'u + a_{12}'v + a_{13}'w = d_1', \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

$$a_{12}'u + a_{22}'v + a_{23}'w = d_2', \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

$$a_{13}'u + a_{23}'v + a_{33}'w = d_3'. \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

These in turn we solve by the escalator process, and having thus found the required values of  $u, v, w$ , we proceed to find the required values of  $x, y, z$ , from (13), (14), (15).

### 3. NUMERICAL EXAMPLE OF APPLICATION OF ESCALATOR PROCESS.

Suppose we wish to solve the following set of six symmetrical linear simultaneous equations, viz.:—

Coefficients of						Constants.
$x.$	$y.$	$z.$	$u.$	$v.$	$w.$	
5399.99	5232.86	4357.85	3622.42	2764.72	1846.91	— 1236.79
5232.86	7871.90	3622.42	5256.51	1846.91	2802.69	— 484.48
4357.85	3622.42	3881.41	2973.04	2639.74	1679.36	— 1249.50
3622.42	5256.51	2973.04	4376.77	1679.36	2632.46	— 473.04
2764.72	1846.91	2639.74	1679.36	2015.78	1149.21	— 1064.70
1846.91	2802.69	1679.36	2632.46	1149.21	1940.65	— 378.31

The procedure may be carried out as follows:—

#### *Stage I.*

$$q_1 = a_{11} = 5399.99,$$

$$x_2 = -a_{12}/q_1 = -0.96904994.$$

We are now in a position to write down the values for rows 1, 2, 3, of Table I. As a check the sum found in row 3, column 1, should be zero.

$$q_2 = a_{12}x_2 + a_{22} = 2800.9973 \text{ is given in row 3, column 2.}$$

If considered desirable, a check on  $q_2$  will be given by the alternative expression for  $q_2$ , viz.,  $q_2 = a_{22} - a_{12}^2/q_1$ .

#### *Stage II.*

We make use of the expressions

$$y_3 = -(a_{13})_2/q_2 = -(a_{13}x_2 + a_{23})/q_2,$$

$$x_3 = -[(a_{13})_1/q_1 + (a_{13})_2x_2/q_2];$$

from which we find  $y_3 = 0.21440731$ ,  $x_3 = -1.01478214$ .

Next rows 4, 5, 6, 7, of Table I. are completed.

TABLE I.

No. of Row.	$r$ .	1.	2.	3.	4.	5.	6.	
1	$a_1x_2$	-5232.8600	-5070.9027	-4222.9743	-3510.3059	-2679.1518	-1789.7480	$d_1x_2$
2	$a_2r$	5232.86	7871.90	3622.42	5256.51	1846.91	2802.69	$d_2$
3	Sum	$(a_{11})_2$	$(a_{12})_2=q_2$	$(a_{13})_2$	$(a_{14})_2$	$(a_{15})_2$	$(a_{16})_2$	$(d_1)_2$
		0	2800.9973	-600.5543	1746.2041	-832.2418	1012.942	714.0313
4	$a_1x_3$	-5479.8134	-5310.2129	-4422.2683	-3675.9671	-2805.5885	-1874.2113	$d_1x_3$
5	$a_2ry_3$	1121.9634	1687.7929	776.6733	1127.0342	395.9910	600.9172	$d_2y_3$
6	$a_3r$	4357.85	3622.42	3881.41	2973.04	2639.74	1679.36	$d_3$
7	Sum	$(a_{11})_3$	$(a_{12})_3$	$(a_{13})_3=q_3$	$(a_{14})_3$	$(a_{15})_3$	$(a_{16})_3$	$(d_1)_3$
		0	0	235.8150	424.1071	230.1425	406.0659	-98.30365
8	$a_1x_4$	9495.1623	9201.2865	7662.6981	6369.5425	4861.3914	3247.5449	$d_1x_4$
9	$a_2ry_4$	-5280.1033	-7942.9691	-3655.1239	-5303.9668	-1863.5843	-2827.9932	$d_2y_4$
10	$a_3r^2_4$	-7837.4790	-6514.8275	-6980.6142	-5346.9345	-4747.5032	-3020.2850	$d_3r^2_4$
11	$a_4r$	3622.422	5256.51	2973.04	4376.77	1679.36	2632.46	$d_4$
		$(a_{11})_4$	$(a_{12})_4$	$(a_{13})_4$	$(a_{14})_4=q_4$	$(a_{15})_4$	$(a_{16})_4$	$(d_1)_4$
12	Sum	0	-0.0001	0	95.4112	-70.3361	31.7267	88.27671



TABLE I. (continued)

No. of Row.	$r$ .	1.	2.	3.	4.	5.	6.		
13	$a_1x_5$	8028.2026	7779.7292	6478.8458	5385.4769	4110.32838	2745.81390	$d_1x_5$	-1838.744264
14	$a_2y_5$	-3432.6045	-5163.7382	-2376.2025	-3448.1182	-1211.51942	-1838.48339	$d_2y_5$	317.804835
15	$a_3z_5$	-10030.7269	-8337.9432	-8934.0761	-6843.2259	-6076.04918	-3865.48446	$d_3z_5$	2876.049707
16	$a_4u_5$	2670.4087	3875.0421	2191.6928	3226.5073	1238.00595	1940.62091	$d_4u_5$	-348.719948
17	$a_5r$	2764.72	1846.91	2639.74	1679.36	2015.78	1149.21	$d_5$	-1064.70
18	Sum	$(a_{11})_5$ -0.0001	$(a_{12})_5$ -0.0001	$(a_{13})_5$ 0	$(a_{14})_5$ 0.0001	$(a_{15})_5=q_5$ 76.54573	$(a_{16})_5$ 131.67696	$(d_1)_5$ 58.309670	
19	$a_1x_6$	-7486.2708	-7254.5703	-6041.5010	-5021.9384	-3832.8668	-2560.46185	$d_1x_6$	1714.622592
20	$a_2y_6$	3836.2951	5771.0184	2655.6553	3853.6333	1353.9999	2054.69781	$d_2y_6$	-355.180200
21	$a_3z_6$	12357.3355	10271.9137	11006.3186	8430.4996	7485.3776	4762.07646	$d_3z_6$	-3543.144138
22	$a_4u_6$	-5798.2901	-8413.9250	-4758.8487	-7005.7537	-2688.0971	-4213.69330	$d_4u_6$	757.17978
23	$a_5r$	-4755.9795	-3177.1269	-4540.9840	-2888.9008	-3467.6236	-1976.91601	$d_5v_6$	1831.538602
24	$a_6r$	1846.91	2802.69	1679.36	2632.46	1149.21	1940.65	$d_6$	-378.31
25	Sum	$(a_{11})_6$ 0.0002	$(a_{12})_6$ -0.0001	$(a_{13})_6$ 0.0002	$(a_{14})_6$ 0.0002	$(a_{15})_6$ 0	$(a_{16})_6=q_6$ 6.35311	$(d_1)_6$ 26.706636	

As checks the values given in row 7, columns 1, 2, should be zero. Column 3 of this row will give

$$q_3 = a_{13}x_3 + a_{23}y_3 + a_{33} = 235.8150,$$

which may be checked by the alternative expression

$$q_3 = a_{33} - (a_{13})_1^2/q_1 - (a_{13})_2^2/q_2.$$

### Stage III.

We make use of the expressions

$$z_4 = -(a_{14})_3/q_3 = -(a_{14}x_3 + a_{24}y_3 + a_{34})/q_3,$$

$$y_4 = -[(a_{14})_2/q_2 + (a_{14})_3/q_3],$$

$$x_4 = -[(a_{14})_1/q_1 + (a_{14})_2x_2/q_2 + (a_{14})_3x_3/q_3];$$

from which we find  $z_4 = -1.79847380$ ,  $y_4 = -1.0090282$ ,  $x_4 = 1.75836664$ .

Next rows 8, 9, 10, 11, 12, of Table I. are completed.

As checks the values given in row 12, columns 1, 2, 3, should be zero. Column 4 of this row will give

$$q_4 = a_{14}x_4 + a_{24}y_4 + a_{34}z_4 + a_{44} = 95.4112,$$

which may be checked by the alternative expression

$$q_4 = a_{44} - (a_{14})_1^2/q_1 - (a_{14})_2^2/q_2 - (a_{14})_3^2/q_3.$$

### Stage IV.

We make use of the expressions

$$u_5 = -(a_{15})_4/q_4 = -(a_{15}x_4 + a_{25}y_4 + a_{35}z_4 + a_{45})/q_4,$$

$$z_5 = -[(a_{15})_3/q_3 + (a_{15})_4z_4/q_4],$$

$$y_5 = -[(a_{15})_2/q_2 + (a_{15})_3y_3/q_3 + (a_{15})_4y_4/q_4],$$

$$x_5 = -[(a_{15})_1/q_1 + (a_{15})_2x_2/q_2 + (a_{15})_3x_3/q_3 + (a_{15})_4x_4/q_4];$$

from which we find  $u_5 = 0.73718914$ ,  $z_5 = -2.30176047$ ,  $y_5 = -0.65597101$ ,  $x_5 = 1.48670693$ .

Next rows 13, 14, 15, 16, 17, 18, of Table I. are completed.

As checks the values given in row 18, columns 1, 2, 3, 4, should be zero. Column 5 of this row will give

$$q_5 = a_{15}x_5 + a_{25}y_5 + a_{35}z_5 + a_{45}u_5 + a_{55} = 76.54573,$$

which may be checked by the alternative expression

$$q_5 = a_{55} - (a_{15})_1^2/q_1 - (a_{15})_2^2/q_2 - (a_{15})_3^2/q_3 - (a_{15})_4^2/q_4.$$

### Stage V.

We make use of the expressions

$$v_6 = -(a_{16})_5/q_5$$

$$= -(a_{16}x_5 + a_{26}y_5 + a_{36}z_5 + a_{46}u_5 + a_{56})/q_5,$$

$$u_6 = -[(a_{16})_4/q_4 + (a_{16})_5u_5/q_5],$$

$$z_6 = -[(a_{16})_3/q_3 + (a_{16})_4z_4/q_4 + (a_{16})_5z_5/q_5],$$

$$y_6 = -[(a_{16})_2/q_2 + (a_{16})_3y_3/q_3 + (a_{16})_4y_4/q_4 + (a_{16})_5y_5/q_5],$$

$$x_6 = -[(a_{16})_1/q_1 + (a_{16})_2x_2/q_2 + (a_{16})_3x_3/q_3 \\ + (a_{16})_4x_4/q_4 + (a_{16})_5x_5/q_5];$$

from which we find  $v_6 = -1.72023913$ ,  $u_6 = -1.60066755$ ,  $z_6 = 2.83564957$ ,  $y_6 = 0.73311632$ ,  $x_6 = -1.38634901$ .

Next rows 19, 20, 21, 22, 23, 24, 25, of Table I. are completed.

As checks the values given in row 25, columns 1, 2, 3, 4, 5, should be zero. Column 6 of this row will give

$$q_6 = a_{16}x_6 + a_{26}y_6 + a_{36}z_6 + a_{46}u_6 + a_{56}v_6 + a_{66} = 6.35311,$$

which may be checked by the alternative expression

$$q_6 = a_{66} - (a_{16})_1^2/q_1 - (a_{16})_2^2/q_2 - (a_{16})_3^2/q_3 - (a_{16})_4^2/q_4 - (a_{16})_5^2/q_5.$$

TABLE II.

Equation.	Substitution of solutions in L.H.S.	Constants of equations.
No. 1	-1236.78970	-1236.79
No. 2	-484.48008	-484.48
No. 3	-1249.49973	-1249.50
No. 4	-473.04014	-473.04
No. 5	-1064.70033	-1064.70
No. 6	-378.31010	-378.31

### Stage VI.

The solutions of the equations are thus

$$\bar{w}_6 = (d_1)_6/q_6,$$

$$\bar{v}_6 = (d_1)_5/q_5 + (d_1)_6v_6/q_6,$$

$$\bar{u}_6 = (d_1)_4/q_4 + (d_1)_5u_5/q_5 + (d_1)_6u_6/q_6,$$

$$\bar{z}_6 = (d_1)_3/q_3 + (d_1)_4z_4/q_4 + (d_1)_5z_5/q_5 + (d_1)_6z_6/q_6,$$

$$\bar{y}_6 = (d_1)_2/q_2 + (d_1)_3y_3/q_3 + (d_1)_4y_4/q_4 + (d_1)_5y_5/q_5 + (d_1)_6y_6/q_6,$$

$$\bar{x}_6 = (d_1)_1/q_1 + (d_1)_2x_2/q_2 + (d_1)_3x_3/q_3 \\ + (d_1)_4x_4/q_4 + (d_1)_5x_5/q_5 + (d_1)_6x_6/q_6;$$

Inserting the values, we obtain

$$\bar{w}_6 = 4.2037106, \quad \bar{v}_6 = -7.99315004,$$

$$\bar{u}_6 = -6.36508248, \quad \bar{z}_6 = 11.59278669,$$

$$\bar{y}_6 = 2.81346708, \quad \bar{x}_6 = -5.38648142.$$

The degree of accuracy achieved is illustrated by Table II., which is extremely satisfactory since the equations are very ill-conditioned in

regard to the ordinary methods of solution ; and the number of significant figures used in the computation, which was done with the aid of a calculating machine, was eight, the coefficients being given to six significant figures.

### *References.*

- (1) J. Morris, "A Successive Approximation Process for solving Simultaneous Linear Equations," Aer. Res. Cttee. Report and Memorandum No. 1711 (1936).
- (2) J. Morris and J. W. Head, "The Escalator Process for the Solution of Lagrangian Frequency Equations," Phil. Mag. ser. 7, xxxv, p. 735 (1944).
- (3) Frazer, Duncan and Collar, 'Elementary Matrices,' Cambridge University Press.

## *XVII. Viscosities of the Di-hydroxy Benzenes and some of their Derivatives.*

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES \*.

[Received May 3, 1945.]

IN phenolic liquids hydrogen bonds of the type  $\text{—O—H} \dots \text{O—}$  may occur both intermolecularly and intramolecularly. It has been demonstrated <sup>(1)</sup> that, in the former case, association causes an elevation of the boiling point with increase in density  $D$ , viscosity  $\eta$ , molar latent heat  $L$ , and rheochor  $R$ . It seems probable that hydrogen bonds of all grades of strength can exist, so that a sharp distinction between strong and weak bonds may not always be possible. Strong bonds yield pronounced absorption in the infra-red <sup>(2)</sup>, a peak occurring at  $c. 7000 \text{ cm.}^{-1}$  in cases of intermolecular association such as that shown by phenol. Such association exerts a great effect on viscosity, and it appears that the presence of weak intermolecular hydrogen bonds may be detected in this way even when they are too weak to give a definite infra-red absorption.

In contrast, intramolecular hydrogen bonding opposes association, and results in a depression of the boiling point and decrease in  $D$ ,  $\eta$  and  $L$ .

The present research embodies a study of the dihydroxy benzenes, in which possibilities of hydrogen bonding occur, together with their methyl and dimethyl ethers.

The *dimethyl ethers* are incapable of yielding hydrogen bonds, and hence follow the established rules, namely that the ortho derivative has higher  $D$ ,  $\eta$  and  $L$  than the meta and para, which latter are closely

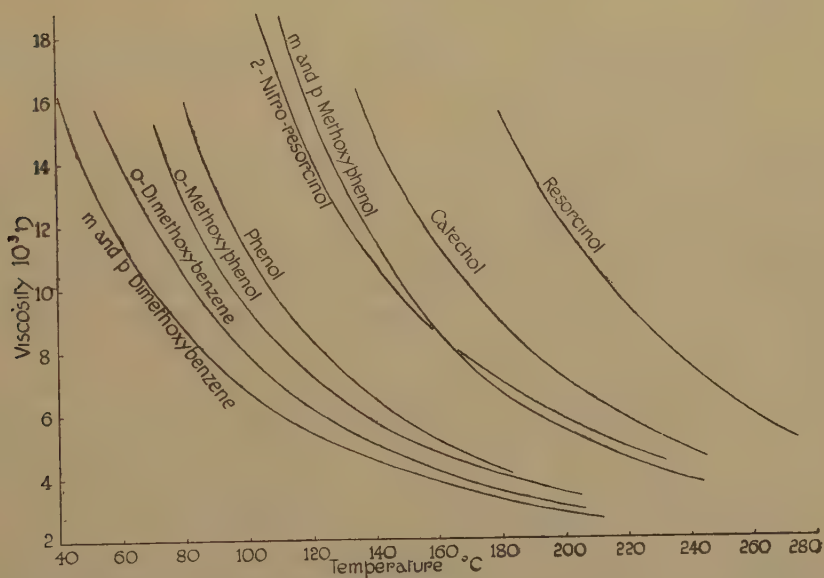
\* Communicated by the Author,



similar (see Table and fig. 1). As there is virtually no association their R/T curves (fig. 2) are practically horizontal.

	B. pt. °C.	B $\propto$ L.	D 150° C.	10 <sup>3</sup> $\eta$ 150° C.	R B. pt. obs.	R* B. pt. calc.
<i>o</i> -dimethoxybenzene ..	206	717	0.9539	4.56	175.8	
<i>m</i> -dimethoxybenzene ..	212	664	0.9482	4.16	175.4	
<i>p</i> -dimethoxybenzene ..	211.6	659	0.9476	4.12	175.1	
<i>o</i> -methoxyphenol .....	205	c. 700	1.0048	5.22	151.3	152.0
<i>m</i> -methoxyphenol ....	245	947	1.0337	9.42	153.0	151.6
<i>p</i> -methoxyphenol ....	244	928	1.0389	9.79	153.4	151.3
<i>o</i> -dihydroxybenzene (catechol) .....	245	1044	1.1202	13.1	129.7	128.2
<i>m</i> -dihydroxybenzene (resorcinol) .....	276	1239	1.1506	30.8	129.5	127.8
2-nitro-resorcinol .....	232	877	1.2902	9.58	156.1	161.2

Fig. 1.



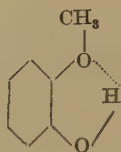
Viscosity temperature curves.

Curves for *m*- and *p*-dimethoxybenzene and *m*- and *p*-methoxyphenol lie too close to be conveniently drawn separately.

The *methoxyphenols* can exert one hydrogen bond. The *meta* and *para* isomerides can associate intermolecularly. The infra-red absorption

\* Using the value for  $C_6H_6$  as observed in the corresponding dimethoxy benzenes.

spectrum <sup>(3)</sup> of the ortho in dilute carbon tetrachloride solution shows a peak at *c.* 6930  $\text{cm}^{-1}$ , which is taken to indicate the presence of a *cis*-form :



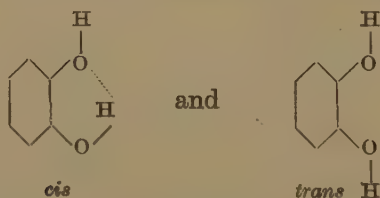
The area of absorption, however, is only about half that observed with phenol (at *c.* 7000  $\text{cm}^{-1}$ ), which suggests that only about 50 per cent. of the molecules are intramolecularly bonded. Intermolecular association would thus be possible with the other molecules, but might be too weak to give definite absorption in the 7000  $\text{cm}^{-1}$  region. This receives support from the following facts :—

(1) Its viscosity is consistently less than that of phenol, but greater than that of the dimethoxybenzenes despite the differences in molecular weights (fig. 1).

(2) Its  $D$ ,  $\eta$  and  $L$  are lower than those of its meta and para isomerides, which latter are closely similar and cannot yield intramolecular bonds (see Table).

(3) The  $R/T$  curves (fig. 2) show that the para isomeride rapidly breaks down to simpler molecules \*, whereas the ortho compound is almost monomeric from 140° C. upwards.

*Catechol and Resorcinol.*—The infra-red absorption spectrum of catechol in dilute carbon tetrachloride solution shows two peaks of almost equal area, one at *c.* 7050  $\text{cm}^{-1}$  as in phenol, and one at *c.* 6960  $\text{cm}^{-1}$ , attributable, as in ortho methoxyphenol, to a *cis*-form. The dipole moment <sup>(4)</sup>,  $2.62 \times 10^{-18}$ , in benzene solution, however, indicates an essentially *cis*-planar structure. There would thus appear to be equilibrium between

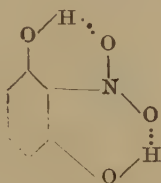


the *cis* predominating.

This *cis*-form has one free hydroxyl and this resembles *m*- (and *p*-) methoxyphenol, whereas the *trans* has two free hydroxyls like resorcinol. A mixture of the two forms should thus yield values for  $\eta$ ,  $D$  and  $L$  intermediate between those of resorcinol and *m*-methoxyphenol.

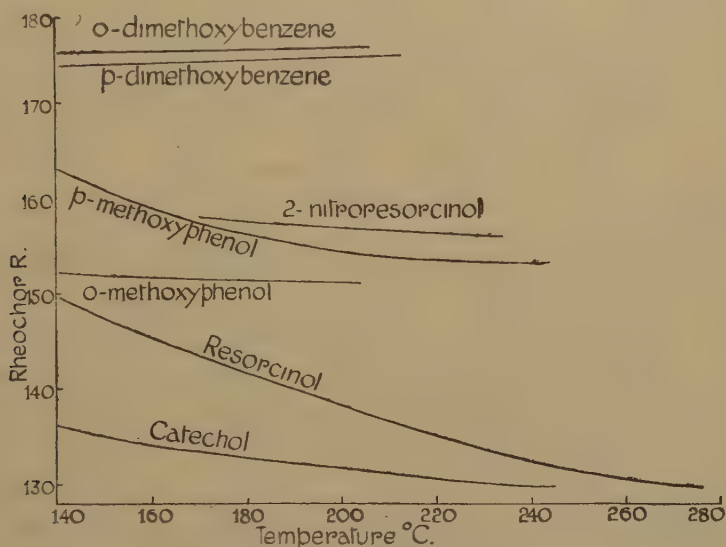
That such is the case is evident from the data in the Table, and the curves in fig. 1. The  $R/T$  curves (fig. 2) show a rapid break-down of association for resorcinol with rise of temperature and a less steep fall for catechol, which is expected as it is less associated to begin with.

*2-nitroresorcinol*.—The infra-red absorption spectrum in dilute carbon tetrachloride solution shows no absorption at *c.* 7000  $\text{cm}^{-1}$ . This suggests complete intramolecular bonding, thus :



and our results support this view.

Fig. 2.



Rheochor temperature curves.

(1) Its viscosity is less than that of resorcinol despite its larger molecular weight and the presence of the  $\text{NO}_2$  group which normally greatly increases viscosity.

(2) The B. pt. and L. are lower than for resorcinol, and the R/T curve (fig. 2) is almost horizontal. At the B. pt. the rheochor is 5.1 units below that calculated, assuming no bridge-ring formation. The decrease is approximately double the value (2.2) found with *o*-nitrophenol, which has only one bridge ring.

### Experimental.

Viscosities were determined with viscometers of the Ostwald type as already outlined <sup>(5)</sup>.

With few exceptions the chemicals were obtained from the British Drug Houses, dried and fractionated.

*p*-dimethoxybenzene was prepared from hydroquinone and dimethyl sulphate <sup>(6)</sup>; it was purified by both distillation and freezing out, until the M. pt. was constant.

2-nitroresorcinol was prepared by Kaufmann's method <sup>(7)</sup>, steam distilled twice and crystallized to constant M. pt. from dilute alcohol.

The viscosity and density data are given below; as before, the densities refer to water at 4° C. The values for A and B refer to Sheppard's equation,  $\log_{10} 10^3 \eta = A + B/T$ , and apply to the linear part of the curve on plotting  $\log 10^3 \eta$  against  $1/T$ . The boiling-point values are given in parentheses, and have been obtained by extrapolation.

*o*-dimethoxybenzene.—B. pt. 206.1° C. at 757.2 mm.  $A = -1.0365$ ;  $B = 717.23$ , from 130° C. upwards. Density at 25° C. 1.0819; Perkin <sup>(8)</sup> found 1.0810.

Temp. ....	53.2	72.5	92.8	112.2	131.0	151.4
D .....	1.0539	1.0348	1.0143	0.9943	0.9743	0.9523
1000 $\eta$ .....	16.47	11.63	8.69	6.76	5.485	4.495

Temp. ....	161.0	171.0	181.3	195.1	206.1	
D .....	0.9418	0.9305	0.9186	0.9027	(0.8900)	
1000 $\eta$ .....	4.131	3.788	3.483	3.133	(2.887)	

$R = 175.8$ .

*m*-dimethoxybenzene.—B. pt. 212.0° C. at 757.7 mm.  $A = -0.9501$ ;  $B = 663.81$  above, 100° C.

Temp. ....	40.5	60.4	80.0	100.3	120.0	139.7
D .....	1.0493	1.0320	1.0144	0.9954	0.9769	0.9582
1000 $\eta$ .....	16.23	11.46	8.676	6.731	5.476	4.560

Temp. ....	160.0	170.2	180.0	190.1	201.0	212.0
D .....	0.9384	0.9277	0.9169	0.9058	0.8937	(0.8815)
1000 $\eta$ .....	3.827	3.521	3.271	3.039	2.818	(2.622)

$R = 175.4$ .

*p*-dimethoxybenzene.—M. pt. 56.3° C. (Mühlhäuser <sup>(9)</sup>, 56.0° C.); B. pt. 211.6° C. at 752.5 mm. D at 55° 1.0383 (Perkin, 1.0375).  $A = -0.9423$ ;  $B = 658.81$ , above 80° C.

Temp. ....	61.8	83.7	102.5	121.9	141.6	159.5
D .....	1.0322	1.0119	0.9939	0.9765	0.9561	0.9382
1000 $\eta$ .....	10.87	8.044	6.296	5.317	4.430	3.811

Temp. ....	170.1	180.3	190.0	199.8	211.6	
D .....	0.9274	0.9164	0.9061	0.8955	(0.8828)	
1000 $\eta$ .....	3.500	3.242	3.024	2.825	(2.613)	

$R = 175.1$ .

*o*-methoxyphenol.—M. pt. 28.3° C. (Perkin <sup>(8)</sup>, 28.3° C.); B. pt. 205.0° C. at 761.1 mm. Sheppard's equation gives a curve. D at 140° 1.1142 (Perkin 1.1143).

Temp. ....	47.2	72.0	93.2	111.6	132.5	140.2
D .....	1.1076	1.0835	1.0625	1.0440	1.0228	1.0149
1000 $\eta$ .....	26.17	15.02	10.37	8.034	6.256	5.754

Temp. ....	150.0	160.1	169.2	179.9	191.2	205.0
D .....	1.0048	0.9947	0.9851	0.9738	0.9619	(0.9474)
1000 $\eta$ .....	5.222	4.727	4.361	3.996	3.670	(3.325)

$= 151.25$ .



*m*-methoxyphenol.—B. pt. 245.2° C. at 749.9 mm. A=−1.2787; B=947.38, above 180° C.

Temp. ....	56.05	80.2	99.9	120.0	139.5	159.6
D .....	1.1151	1.0948	1.0780	1.0604	1.0432	1.0252
1000 $\eta$ ....	75.87	34.77	21.77	14.76	10.89	8.279
Temp. ....	179.8	195.0	209.2	220.1	230.0	245.2
D .....	1.0063	0.9923	0.9788	0.9684	0.9590	(0.9446)
1000 $\eta$ .....	6.506	5.562	4.856	4.390	4.024	(3.546)

R=153.0.

*p*-methoxyphenol.—M. pt. 52.5° C. (Hlasiwetz <sup>(10)</sup>, 53.0° C.); B. pt. 244.0° C. at 754.2 mm. A=−1.2227; B=927.99, above 170° C.

Temp. ....	72·3	92·0	112·0	131·5	150·9	172·6
D .....	1·1088	1·0927	1·0737	1·0561	1·0382	1·0179
1000 $\eta$ .....	45·80	27·45	18·13	12·96	9·782	7·260
Temp. ....	183·0	195·0	208·2	221·5	234·0	244·0
D .....	1·0078	0·9963	0·9833	0·9706	0·9583	(0·9484)
1000 $\eta$ .....	6·489	5·713	5·055	4·495	4·016	(3·733)
R=	153·4.					

*o*-dihydroxybenzene (Catechol).—M. pt. 105° C. (International Critical Tables <sup>(11)</sup>, 105° C.; Fittig <sup>(12)</sup>, 104° C.); B. pt. 245° C. at 749.8 mm. A=−1.3560; B=1043.5, above 160° C.

Temp. ....	121·0	140·7	160·0	170·2	179·3	190·0
D .....	1·1493	1·1297	1·1102	1·0999	1·0902	1·0793
1000 $\eta$ .....	21·71	15·01	11·351	9·954	8·932	7·898
Temp. ....	200·7	210·1	220·6	232·0	245·0	
D .....	1·0680	1·0584	1·0477	1·0353	(1·0212)	
1000 $\eta$ .....	7·024	6·368	5·794	5·152	(4·554)	
R=129·7.						

*m*-dihydroxybenzene (Resorcinol).—M. pt. 110° C. (Philip <sup>(13)</sup>, 110.5° C.); B. pt. 276° C. at 761.2 mm. A = -1.5444; B = 1238.7, above 180° C.

Temp. ....	141.0	159.0	183.0	198.3	220.0
D .....	1.1586	1.1430	1.1217	1.1079	1.0890
1000 $\eta$ .....	37.55	24.93	14.92	12.14	9.348
Temp. ....	231.2	242.0	250.1	260.0	276.0
D .....	1.0792	1.0693	1.0624	1.0535	(1.0391)
1000 $\eta$ .....	8.168	7.296	6.668	6.012	(5.158)
R=129.5.					

*2-nitroresorcinol*.—M. pt. 85° C. (International Critical Tables <sup>(11)</sup>, 85°). B. pt. taken as 232° C. Decomposition occurred with marked darkening above 180°.  $A = -1.0924$ ;  $B = 877.43$ , above 150° C.

Temp. ....	96.7	119.6	139.4	150.1
D .....	1.3526	1.3258	1.3026	1.2901
1000 $\eta$ .....	21.31	14.40	10.97	9.581
Temp. ....	159.7	170.0	180.5	232.0
D .....	1.2783	1.2663	1.2531	(1.1884)
1000 $\eta$ .....	8.604	7.708	6.952	(4.415)

R=156.1.

## References.

- (1) Friend and Hargreaves, *Phil. Mag.* (In the Press.)
- (2) Wulf and Liddel, *J. Amer. Chem. Soc.* lvii. p. 1464 (1935); Pauling, *ibid.* lviii. p. 94 (1936); also 'Nature of the Chemical Bond' (1939).
- (3) Wulf, Liddel and Hendricks, *J. Amer. Chem. Soc.* lviii. p. 2287 (1936).
- (4) Lander and Svirebely, *J. Amer. Chem. Soc.* lxvii. p. 322 (1945).
- (5) Friend and Hargreaves, *Phil. Mag.* xxxiv. p. 643 (1943); Hargreaves, *ibid.*
- (6) Ullmann, *Annalen*, cccxxvii. p. 116 (1903).
- (7) Kaufmann and de Pay, *Ber.* xxxvii. p. 725 (1904).
- (8) Perkin, *J. Chem. Soc.* lxix. p. 1188 (1896).
- (9) Mühlhäuser, *Annalen*, ccvii. p. 252 (1881).
- (10) Hlasiwetz and Habermann, *Annalen*, clxxvii. p. 340 (1875).
- (11) "International Critical Tables," i. pp. 198-199.
- (12) Fittig and Mager, *Ber.* viii. p. 365 (1875).
- (13) Philip and Smith, *J. Chem. Soc.* lxxxvii. p. 1737 (1905).

The Technical College,  
Birmingham.  
May 1, 1945.

XVIII. *The Complete, Three-Dimensional Colour Domain and its Metrics.*

By LUDWIK SILBERSTEIN, PH.D.\*

Communication No. 1027 from the Kodak Research Laboratories.

[Received June 8, 1945.]

THE domain of all colours, as perceived by human eyes, is a manifold. As we have explained more fully elsewhere †, a manifold, whether discrete or continuous, has by itself neither metrical properties nor even dimensionality. To give it the latter, its elements, in this case colours, must first be ordered in some way, and since this can be accomplished in a variety of ways, the dimensionality of the manifold will differ according to the ordering principle chosen. This is palpably true of a discrete manifold, but it holds equally well for any continuous or supposedly continuous manifold, such as dreary space and space time or the fascinating domain of all colours. As a matter of fact, experimentally, the question whether a manifold *is* or is not continuous is devoid of meaning. It is only sensible to ask whether, with our actual mathematical tools and the degree of accuracy of our art of experimenting and observing, it is more convenient and fruitful to *treat* a manifold as continuous rather than discrete.

So far as we are able to judge, it is certainly most convenient and expedient to treat the colour domain as a *continuous* manifold. Of how many dimensions?

\* Communicated by Kodak, Ltd.

† 'Theory of Relativity,' 2nd ed. Macmillan, London, ch. xii. (1924).

It will be obvious to all conversant with the subject that ordering principles resembling those used in school geometry, where (say) a line is conceived as generated by the motion of a point, a surface as generated by the motion of the line, not along itself, a solid as produced by the motion of the surface not in itself, do not readily offer themselves in the case of the domain of colours. Again, of the triad : hue, saturation and luminosity ("Helligkeit"), originally proposed by Helmholtz, the third, at least, is not a well-defined concept and—from the general tensorial viewpoint—has no intrinsic, *i. e.* generally invariant, meaning at all.

On the other hand, it is certainly possible to exhaust the whole colour domain by just *three* independent components, the so-called *stimuli co-ordinates*,  $x, y, z$ , as used by most modern and even some older investigators of the domain, briefly, by most colorimetrists. Such an  $x, y, z$ -system of colour co-ordinates has even been internationally standardized.

We shall, therefore, without more ado, adopt this  $x, y, z$ -system for the present investigation.

The whole colour domain being thus exhaustible and the triad  $x, y, z$  of component intensities (say, red, green, and violet) determining each time a unique visual colour (though not conversely), we shall forthwith treat the complete colour domain as a *three-dimensional* manifold, briefly, a *threefold*. Such, *e. g.*, is also the space of projective geometry. To begin with, both the colour domain and the projective space are *non-metrical* or, as some modern mathematicians say, "*amorphous*."

Our ordering principle, the stimuli-co-ordinates  $x, y, z$ , have made the domain three-dimensional, but have, so far, left it amorphous, non-metrical.

Metrics have yet to be *impressed* upon this amorphous manifold or class of elements, colours, which are, so far, merely labelled by triple labels,  $x, y, z$ . As stated earlier, this can be done in a variety of ways. To give the domain metrics, colorimetrics, we have to decide what metricizing principle is to be adopted. In this choice we should be guided by considerations of convenience and heuristic value, *viz.*, leading to a discovery of some new and interesting peculiarities of the colour domain.

The choice of a metricizing principle is, technically, equivalent to impressing upon the amorphous manifold a suitable "*metrical tensor*," *i. e.* an array of magnitudes covariant with respect to *any* transformations whatever of  $x, y, z$  (not just linear ones). For, only such tensorial entities do express some intrinsic properties of the manifold. Our choice of the *colorimetric tensor*, and the reasons for its preference, will be explained presently.

The complete colour domain, having been converted through the introduction of the stimuli-co-ordinates,  $x, y, z$ , into a three-dimensional, but, so far, amorphous, manifold, can have only *non-metrical* tensors, and actually has just one such tensor, a contravariant tensor of rank one or *vector*, namely, the infinitesimal vector pointing from  $P(x, y, z)$  to another colour  $Q(x+dx, y-dy, z+dz)$ , and thus consisting of  $3^1=3$  components,

$$dx, dy, dz$$

or, briefly, with the conventional superscript,

$$dx^{\iota}(\iota=1, 2, 3).$$

This is a genuine tensor\*, in fact, the prototype of all contravariant first-rank tensors. Its transformation rule, in passing from  $x, y, z$  to any other co-ordinates  $x', y', z'$ , is

$$dx'=\frac{\partial x'}{\partial x}dx+\frac{\partial x'}{\partial y}dy+\frac{\partial x'}{\partial z}dz,$$

and similarly for  $dy'$  and  $dz'$ . The new components of the vector are linear homogeneous functions of the old ones, with coefficients  $\partial x'/\partial x$ , etc., varying in general from colour to colour. The three transformation formulæ are usually written, in compact form :

$$dx^{\iota'}=\frac{\partial x^{\iota'}}{\partial x^{\kappa}}dx^{\kappa}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

to be summed over the index (superscript) appearing *twice*, i. e., over  $\kappa=1, 2, 3$ .

It will be recalled that a *tensor* of any rank is an array of magnitudes (generally functions of the co-ordinates) which obey a homogeneous transformation rule for all conceivable co-ordinate transformations whatever.

Such and only such entities have a meaning. All other concoctions, as say,  $x+y+z$  or  $xy, yz, zx$ , do not express anything concerning the manifold, in our case, threefold, since their form depends on the fortuitous selection of the co-ordinate system. Such arrays of magnitudes are not tensors, but—to quote a phrase used sometimes by tensor students—“monsters.” We shall do our best to avoid them.

Conversely, in passing from  $x'$  to  $x$ , we have the transformation rule

$$dx^{\iota}=\frac{\partial x^{\iota}}{\partial x^{\iota'}}dx^{\iota'}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1')$$

Our colour domain has, to begin with, actually only this tensor or vector,  $dx'$ , that is,  $dx^1, dx^2, dx^3$ , or, as we shall frequently write,  $dx, dy, dz$ . So long as no metrics are impressed upon the domain, it has, for the experimentalist or observer, no other tensors whatever, finite or infinitesimal.

We may recall for the sequel that a *tensor of rank zero* is (or consists of) just a *single* magnitude ( $3^0=1$ ), say  $I$ . Its transformation rule is, of course,  $I(x', y', z')=I(x, y, z)$  or, briefly,

$$I'=I.$$

In other words, such a tensor is just a single magnitude, *invariant*, i.e., the same, no matter what the choice of the system of co-ordinates. It may be a mere constant, all over the manifold, or it may vary from “place to place,” in our case, from colour to colour. In this, the more general case, it is necessarily an *invariant function* of the co-ordinates. Needless to say, the sum  $x+y+z$  of the co-ordinates of a colour (sum of component intensities) is not an invariant of the colour domain. It has

---

\* While the finite co-ordinates  $x, y, z$  are *not* the components of a general tensor.



no intrinsic meaning, for the observer, in spite of the fact that it has a physical ("objective") meaning, *i.e.*, that of the *resultant intensity* of the stimulus offered to the observer's eye.

We have said a moment ago that the colour domain has, so far, no other tensor than the infinitesimal vector  $dx, dy, dz$ .

Abstractly, we may speak of other tensors of the domain, as, *e.g.*, of a second-rank covariant tensor  $Q_{ik}$  ( $3^2=9$  components) transformed by the rule

$$Q_{ik}' = \frac{\partial x_\alpha}{\partial x_i'} \frac{\partial x_\beta}{\partial x_k'} Q_{\alpha\beta},$$

to be summed over  $\alpha=1, 2, 3$  and  $\beta=1, 2, 3$ . But in our amorphous colour domain no such tensors, with a definite colour meaning, so to speak, present themselves. (One such tensor will be introduced presently, *viz.*, in connection with actual colour *matching*.) At any rate, however, none of such tensors, whether purely conceptual or experimental, has an invariant of its own. Such also is the case of our tensor (vector)  $dx, dy, dz$ . No such invariants exist.

This negative theorem happens, in our case, to be of some importance, not just mathematically, but also with regard to the interests of the experimental student of the intrinsic properties of the colour domain, always in its subjective aspect. (*E.g.*, luminosity or Helmholtz's "Helligkeit" cannot be said to possess an intrinsic meaning, since it is not a general invariant.) Schrödinger's "Helligkeit," for instance, so elaborately introduced in his paper (*Ann. der Phys.* III. vol. lxxiii. 1920), is simply the sum of the stimuli-co-ordinates  $h=x+y+z$ , and is, therefore, far from being a general invariant. We may have  $h=h'$  only for some very special (not all) linear transformations of the co-ordinates, to say nothing of other, non-linear transformations. In fine, Schrödinger's luminosity is not a tensor at all.

The tensorial point of view rules out, of course, not only Schrödinger's luminosity, the simple sum of  $x, y, z$ , but also any other, more elaborate function of the three co-ordinates. Needless to say, the same essential objection can be raised against the concept of equiluminosity of any two otherwise distinct colours.

This, then, would be the negative part of the results of our present inquiry. Now for its positive or, as we hope, constructive results.

A colour C being produced in one half of the field in a trichromatic colorimeter, we ask the observer to set the three independent knobs  $K_1, K_2, K_3$ , of the instrument so that the other half of the field is (for him, of course) indistinguishable from C extending over the first half (matching of C). The observer does his best and, in a number ( $n$ ) of trials, finds as many matches  $Q(Q_1, Q_2, \dots Q_n)$ , differing more or less physically, *i.e.*, as to their co-ordinates  $x, y, z$ , from each other and from the central colour  $C(x_0, y_0, z_0)$ .

For simplicity let C be the local origin of co-ordinates so that  $x_0=y_0=z_0=0$ .

Now, the probability, say,

$$f(Q)\Delta\tau=f(x, y, z)\Delta\tau,$$

that a match  $Q$  shall fall within some definite little domain  $\Delta\tau$  around  $Q(x, y, z)$ , may be safely (or, at least, tentatively) assumed to be a general invariant function of  $Q(x, y, z)$  and of the sub-domain  $\Delta\tau$  chosen by us around  $Q$ .

Notice that the factor  $f(Q)$  itself need not be an invariant, only its product into  $\Delta\tau$ , the content of the sub-domain, is here assumed to be a general invariant, that is, independent of the choice of the co-ordinate system or, experimentally, remaining essentially the same when the knobs  $K_1, K_2, K_3$  of the instrument (measuring  $x, y, z$ , say, by the angles of rotation from the position corresponding to  $x=y=z=0$ ) are replaced by others,  $K_1', K_2', K_3'$  (with  $x', y', z'$ =angles of *their* rotation), no matter how complicated their mechanism as compared with that of the original knobs. This, in fact, is equivalent to replacing  $x, y, z$ , by  $x', y', z'$ , any functions whatever of  $x, y, z$ .

In particular, the probability that a match shall fall within a domain  $\Delta\tau$  surrounding  $C$  itself,

$$f(0, 0, 0)\Delta\tau,$$

will be such an invariant.

The probability that two matches,  $Q_1$  and  $Q_2$ , shall fall within two domains of content  $\Delta\tau_1$  and  $\Delta\tau_2$  will be

$$P=f(Q_1)f(Q_2) \cdot \Delta\tau_1 \cdot \Delta\tau_2$$

or, passing to logarithms and to any number ( $n$ ) of matches  $Q$ ,

$$\log P=\Sigma\{\log f(Q)+\log \Delta\tau\},$$

the sum to be extended over the  $n$  matches. Let us consider the probability of a single match with a (surrounding) domain of infinitesimal content \*,

$$P=f(x, y, z)d\tau,$$

and let us try to derive some simple conclusions from its invariance. Suppose that, with a certain choice of the co-ordinate system  $x, y, z$ , we note down our element  $d\tau$  as the simple product  $dx \cdot dy \cdot dz$ , so that in these co-ordinates and with  $x, y, z$ , replaced by  $x_1, x_2, x_3$ , the probability in question is  $P=f(x_1, x_2, x_3) \cdot dx_1 dx_2 dx_3$  or, more compactly,

$$P=f(x)dx_1 dx_2 dx_3.$$

As we know, neither  $f(x)$  nor  $dx_1 dx_2 dx_3$  is an invariant, but their product is, by our original assumption, a general invariant. Let now  $x_1', x_2', x_3'$ , be any other system of co-ordinates, functions of the old ones. In these co-ordinates the probability of the contemplated match will be expressed by

$$f(x)dx_1 dx_2 dx_3=J \cdot f(x')dx_1' dx_2' dx_3', \quad . \quad . \quad . \quad . \quad (3)$$

---

\* I. e., such that the variations of the function  $f(x, y, z)$  within this domain are negligible.

where  $J$  is the *Jacobian* of the transformation, the determinant

$$\begin{vmatrix} \partial x_1 / \partial x_1' & \dots & \partial x_1 / \partial x_3' \\ \dots & \dots & \dots \\ \dots & \dots & \partial x_3 / \partial x_3' \end{vmatrix},$$

briefly written

$$\left| \frac{\partial x_i}{\partial x_k'} \right|,$$

which will be assumed not to vanish. Now, as is well known,

$$dx_1' \cdot dx_2' \cdot dx_3' \left| \frac{\partial x_i}{\partial x_k'} \right| = dx_1 dx_2 dx_3,$$

Hence, the function  $f(x)$  or  $f(Q)$ , which may be called the *frequency* of matches at the place  $Q$ , is an *invariant*,

$$f(x') = f(x), \text{ i. e., } f(x', y', z') = f(x_1 y_1 z). \quad . \quad . \quad . \quad (4)$$

This important property may be written thus

$$f'(Q) = f(Q).$$

This holds for any  $Q$ . In particular, if  $Q$  coincides with the colour centre, we have

$$f'(0) = f(0) \quad \text{or} \quad f'(C) = f(C),$$

which stand for

$$f'(0, 0, 0) = f(0, 0, 0).$$

Write  $f(x, y, z) = Ag(x, y, z)$ , where  $g$  is some function of  $x, y, z$ , such that

$$g(0, 0, 0) = 1.$$

Then  $A = f(0, 0, 0)$  and our frequency is expressed by

$$f(x, y, z) = f(0, 0, 0) \cdot g(x, y, z), \quad . \quad . \quad . \quad (5)$$

so that also

$$\log f(x, y, z) = \log f(0, 0, 0) + \log g(x, y, z). \quad . \quad . \quad . \quad (5')$$

Since  $f = f'$  an invariant, we have, by (5),

$$f'(0, 0, 0) g(x', y', z') = f(0, 0, 0) g(x, y, z), \quad . \quad . \quad . \quad (6)$$

which holds for any co-ordinate transformations which leave the central co-ordinates  $0, 0, 0$  intact.

The frequency  $f = f(0) \cdot g$  must then, at any rate, be a function of some invariant of the colour domain. But our domain has, so far, no invariant nor any other well-defined tensor than the simple contravariant vector  $dx'$  or  $dx, dy, dz$ , and this vector has, in the amorphous domain, no invariant of its own.

To provide for an invariant associated with this vector, we must *impress* some other tensor upon the colour domain.

The simplest way is to impress upon it a *covariant vector*  $A_k$ , finite or infinitesimal. Then, since our original vector  $dx^k$  is contravariant, the inner product of the two vectors,

$$A_k dx^k = A_1 dx + A_2 dy + A_3 dz,$$

will be an invariant. If  $A_i$  is to be, like  $dx^\kappa$  itself, an infinitesimal (covariant) vector associated somehow with  $dx^\kappa$ , we are necessarily driven to take

$$A_1 = c_{11}dx + c_{12}dy + c_{13}dz, \text{ etc.},$$

or briefly,

$$A_i = c_{i\kappa}dx^\kappa,$$

where the  $c_{i\kappa}$  must be the components of a legitimate second-rank covariant tensor. In this manner our last invariant will become

$$A_\kappa dx^\kappa = c_{i\kappa} dx^i dx^\kappa,$$

or

$$c_{11}dx^2 + (c_{12} + c_{21})dx dy + \dots + c_{33}dz^2.$$

This, then, amounts ultimately to impressing upon the amorphous colour domain a *covariant second-rank tensor*  $c_{i\kappa}$ . Since in the resulting invariant (which is our only objective),  $c_{12}$  and  $c_{21}$  appear only through their sum, and similarly for the remaining two pairs of components, we may as well use a symmetrical tensor, *i. e.*,

$$c_{i\kappa} = c_{\kappa i}.$$

We shall denote the differential invariant thus generated by

$$m \cdot ds^2,$$

where, for future convenience,  $m$  will stand for some pure number, such as 2 or 1/2.

Thus, our invariant will be

$$m ds^2 = c_{i\kappa} dx^i dx^\kappa = c_{11}dx^2 + 2c_{12}dx dy + \dots + c_{33}dz^2. \quad (7)$$

The frequency or matching probability per unit content  $\Delta\tau$  must then be some function of this invariant alone,

$$f = f(ds^2) = f(c_{i\kappa} dx^i dx^\kappa). \quad (8)$$

What kind of function? In the first place we have, as in (5),

$$f(ds^2) = f(0) g(c_{i\kappa} dx^i dx^\kappa), \quad (9)$$

where  $f(0) = f(0, 0, 0)$  is some number attached to the central colour C. Thus,

$$\log f = \log f(0) + \log g(c_{i\kappa} dx^i dx^\kappa), \quad (10)$$

and

$$\log P = \log [f(0)d\tau] + \log g(c_{i\kappa} dx^i dx^\kappa). \quad (10 a)$$

Here,  $\log g$  has to be some function of the invariant argument  $c_{i\kappa} dx^i dx^\kappa$  only. The simplest device is to make it, tentatively at least, a *linear* function of that argument,  $\log g = \text{const. } c_{i\kappa} dx^i dx^\kappa + \text{constant}$ , and since the additive constant may be thrown upon the coefficient  $A = f(0)$  and the numerical multiplicative constant upon each of the six tensor components  $c_{i\kappa}$ , we may as well take, without any loss to generality,

$$\log g = -c_{i\kappa} dx^i dx^\kappa,$$

while we assume, for what is to follow, that the tensor  $c_{i\kappa}$  is such as to make  $c_{i\kappa} dx^i dx^\kappa$  a *definite, positive* quadratic form, that is, always positive, for any vector  $dx^\kappa$ .



Our frequency function now becomes  $f=f(0)e^{-m\delta s^2}$ ,

or 
$$f(x, y, z)=Ae^{-c_{i\kappa}dx^i dx^\kappa}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and the probability of a match falling into a domain  $d\tau=dx dy dz$ ,

$$P=Ae^{-c_{i\kappa}dx^i dx^\kappa} dx dy dz.$$

The formula (11) thus derived, however, is the expression of what, since Bravais' time, has been called *the normal distribution law of frequencies for three attributes*, in the present case, the stimuli-co-ordinates of a colour.

We shall henceforth write  $x, y, z$ , instead of  $dx^i$  ( $i=1, 2, 3$ ) or  $dx, dy, dz$ . It will be kept in mind, however, that  $x, y, z$ , which replace our original co-ordinate differences  $x-x_0$ , etc., are to be, in the practical sense of the word, *infinitesimal, i. e.*, ultimately, small as compared with the total  $x$ -,  $y$ -, or  $z$ -extent of the whole colour domain and, therefore, not to exceed a very few *standard deviations*\* ( $\sigma$ ), in any "direction," of the colour matches. With this understanding, our "normal" frequency function will be written, concisely,

$$f(x, y, z)=Ae^{-c_{i\kappa}x_i x_\kappa},$$

and fully, 
$$f(x, y, z)=Ae^{(c_{11}x^2+2c_{12}xy+\dots+c_{33}z^2)}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Historically, we may mention that this law (in two dimensions, viz., for the distribution of rifle shots on a target) was derived (1772) by Cotes, and later on generalized to any number of "attributes," from the "Principle of the Centroid," an assumption formulated (say, in our case) as follows :

If  $x_1, y_1, z_1; x_2, y_2, z_2; \dots x_n, y_n, z_n$  be  $n$  actual matchings, then the most probable colour (centre) matched by them is their centroid, that is, a colour whose co-ordinates are

$$[x]=\frac{1}{n}\sum_1^n x_\kappa, \quad [y]=\frac{1}{n}\sum_1^n y_\kappa, \quad [z]=\frac{1}{n}\sum_1^n z_\kappa \dagger.$$

The details of the simple proof that the centroid principle leads to the normal frequency distribution, as in (12), will be left to the reader.

Cotes' centroid principle has, no doubt, a good deal of intuitive plausibility and brings out, at the same time, a noteworthy property of the "normal" law.

Here, however, we have arrived at the normal three-attribute (viz., trichromatic) law with a minimum of actual assumptions, by elementary *tensorial* considerations, viz., admitting only such entities which have an *intrinsic* meaning independent of the co-ordinate system.

\* A concept to be developed presently, together with its mathematical expression.

† Throughout this paper,  $[\phi]$  will stand for the "mean" or "the probable" value of  $\phi$ ;  $[\phi]=\frac{1}{n}\sum_1^n \phi_\kappa$ . More generally, if  $p_1, p_2, \dots p_n$  are the probabilities of  $\phi=\phi_1, \phi_2$ , etc.,  $\phi_n$ , then  $[\phi]=\frac{1}{n}\sum_1^n p_\kappa \phi_\kappa$ .

There is more, however, to recommend this simple, yet powerful, modern method. Namely, it enables us, without any integrations over  $x, y, z$ , to evaluate the factor  $A$  in the formula (12) belonging to any colour centre  $C$ , apart from a purely numerical, universal multiplier.

Since  $c_{ik}$  is a tensor, its determinant  $D = |c_{ik}|$ , without being an invariant, has the remarkable property

$$D' = J^2 \cdot D, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where

$$J = \left| \frac{\partial x_i}{\partial x'_k} \right|$$

is the Jacobian already mentioned. On the other hand, as is well known,

$$dx dy dz = dx' dy' dz' \cdot J.$$

Now, since

$$P = A e^{-c_{ik} x_i x_k} dx dy dz$$

as well as  $e^{-c_{ik} x_i x_k}$  are invariant, the product  $A dx dy dz$  is an invariant. And since

$$\sqrt{D} \cdot dx dy dz = dv \text{ (say)}$$

is an invariant, we may write

$$P = \frac{A}{\sqrt{D}} e^{-c_{ik} x_i x_k} dv,$$

and we now see that, since the second and the third factors are invariant,

$\frac{A}{\sqrt{D}}$  is an invariant number, say  $N$ , and  $A = N\sqrt{D}$ .

Ultimately, therefore, we have

$$P = N\sqrt{D} \cdot e^{-c_{ik} x_i x_k} dv, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where  $D = |c_{ik}|$ .

This simple result agrees with the value of the coefficient  $A$  obtainable in a quite different way. Namely, the probability that a match of a colour  $C$  shall fall anywhere within the whole colour domain or, practically, within a moderately small\* domain  $v$  around  $C$ , is *unity*, that is to say,

$$A \iiint_{(v)} e^{-c_{ik} x_i x_k} dx dy dz \doteq A \iiint_{-\infty}^{\infty} e^{-c_{ik} x_i x_k} dx dy dz = 1.$$

Now, the integral is found, after some rather laborious but straightforward manipulations, to be equal to  $\sqrt{\pi^3}/D$ . Consequently,

$$A = \sqrt{\frac{D}{\pi^3}},$$

proportional to the square-root of  $D = |c_{ik}|$ , in perfect agreement with our result based only on the invariance of the probability  $P$  and on

---

\* A domain, that is, whose  $x$ -,  $y$ -,  $z$ -dimensions amount only to a few "standard deviations."

tensorial considerations of the simplest kind. The number  $N$  or  $A/\sqrt{D}$  is here  $\pi^{-3/2}$  and thus certainly an invariant, viz., a universal constant, a pure number as good as  $\pi$  itself.

We have in this manner derived the trichromatic normal law, for the complete, three-dimensional colour domain. Our derivation was based, substantially, on the assumption that the matching probability  $P$  is an *invariant* with respect to the selection of the labelling system  $(x, y, z)$  of colours or independent of the structure of the three knobs of the colorimeter—each to be handled freely by the “observer” or colour matcher. The normal law is a logical implication of this assumption and of the deliberate choice of the simplest *form* of  $\log g$  function of the invariant  $c_{i\kappa} x_i x_\kappa$ .

The hypothetical invariance of the probability  $P$  can, at any rate, be tested. We have only to replace our  $K_1, K_2, K_3$  by some other knobs,  $K'_1, K'_2, K'_3$ , and to ascertain whether the matches around the chosen colour  $C$  have, essentially, the same distribution, *e.g.*, the same fan of standard deviations as with  $K_1, K_2, K_3$ . This is worth trying directly.

But a simpler way of testing the legitimacy of all our assumptions upon which the normal trichromatic law for the probability  $P=f(x, y, z)dv$  or the corresponding frequency,

$$f(x, y, z) = \sqrt{\frac{D}{\pi^3}} e^{-c_{i\kappa} x_i x_\kappa} \quad . \quad . \quad . \quad . \quad . \quad (15)$$

is based, is to test experimentally all possible implications of such a frequency distribution. Such tests may, in fact, be considerably easier and, perhaps, also much more interesting than a direct test of our principal assumption, the invariance of the matching probabilities.

We shall now consider in some detail such implications of the trichromatic frequency distribution (15), *i.e.*, derive from it a set of formulae which may be needed for the application of this frequency function to actual colorimetric experiments or observations.

The first question of this kind, which naturally suggests itself, is as follows :—

A number  $(n)$  of free trichromatic matchings of a given colour  $C$  having been made, what are *the most probable* values of the six components of our tensor  $c_{i\kappa}$  (and hence of the factor  $A=\sqrt{|c_{i\kappa}|/\pi^3}$ ), that is to say, values for which the set of those matchings  $(x, y, z)_1 \dots (x, y, z)_n$  as a compound event is most probable ?

To answer this question, it is enough to write down the first necessary condition for a maximum of the product  $f_1, f_2 \dots f_n$ , where  $f_\kappa=f(x_\kappa, y_\kappa, z_\kappa)$ , or of the sum  $\sum_{\kappa=1}^n \log f_\kappa$ , when the  $c_{i\kappa}$  are varied. This condition splits into the six equations

$$\frac{\partial}{\partial c_{i\kappa}} \sum \log f_\kappa = 0.$$

If we introduce the abbreviations,

$$a, b, c = [x^2, y^2, z^2],$$

$$d, e, f = [yz, zx, xy],$$

$$\lambda, \mu, \nu = 1 + \frac{1}{4} \left( \frac{bc}{d^2}, \frac{ca}{e^2}, \frac{ab}{f^2} \right),$$

and

$$R = 1 + \frac{8(2ef - ad)(2fd - be)(2de - cf)}{(4d^2 + bc)(4e^2 + ca)(4f^2 + ab)},$$

then the said six equations assume the comparatively simple form

$$\left. \begin{aligned} c_{11}, c_{22}, c_{33} &= \frac{1}{2R} \left\{ \frac{1}{a} \left( R - \frac{1}{\mu} - \frac{1}{\nu} \right), \frac{1}{b} \left( R - \frac{1}{\nu} - \frac{1}{\lambda} \right), \frac{1}{c} \left( R - \frac{1}{\lambda} - \frac{1}{\mu} \right) \right\}, \\ c_{23}, c_{31}, c_{12} &= \frac{1}{4R} \left\{ \frac{1}{d\lambda}, \frac{1}{e\mu}, \frac{1}{f\nu} \right\}. \end{aligned} \right\} \quad \dots \quad (16)$$

These are—computationally—quite simple expressions for all required coefficients or tensor components in terms of the averages  $a = [x^2]$ , etc., extended over all matches of the chosen central colour.

If, e. g., the matchings are such that  $d = [yz]$ ,  $e = [zx]$ ,  $f = [xy]$  all vanish, then

$$c_{12} = c_{23} = c_{31} = 0$$

and  $\lambda = \mu = \nu = \infty$ , so that, by the first line of (16),

$$c_{11}, c_{22}, c_{33} = \frac{1}{2} [x^2, y^2, z^2], \quad \dots \quad (16a)$$

at or around the given colour C, of course. Thus, the exponent in the frequency function becomes

$$c_{ik} x_i x_k = c_{11} x^2 + c_{22} y^2 + c_{33} z^2.$$

In other words, if  $[yz] = [zx] = [xy] = 0$ , then only the diagonal components  $c_{kk}$  of the tensor survive. The axes of  $x, y, z$  are in this case (always at C only) the *principal axes* of the tensor  $c_{ik}$ . At the same time,  $D = c_{11} \cdot c_{22} \cdot c_{33}$ , so that the frequency function becomes

$$f(x, y, z) = \sqrt{\frac{c_{11} c_{22} c_{33}}{\pi^3}} e^{-c_{11} x^2} \cdot e^{-c_{22} y^2} \cdot e^{-c_{33} z^2}, \quad \dots \quad (15a)$$

a mere product of three one-dimensional normal or Gaussian distribution functions. It goes without saying that if the original co-ordinate axes are devoid of this remarkable property, we can always introduce (locally) a system of axes,  $X', Y', Z'$ , which converts  $c_{ik} x_i x_k$  into

$$c_{11}' x'^2 + c_{22}' y'^2 + c_{33}' z'^2,$$

namely, by a rigid rotation of the original system of axes. The co-ordinates  $x', y', z'$  are then linear homogeneous functions of  $x, y, z$ . This can be accomplished separately for every colour chosen as the centre of matchings and, thus, throughout the colour domain.



The *principal axes* of the tensor  $c_{ik}$  which is to become our *metrical tensor* are, of course, intrinsically remarkable "directions" at any point (colour) of the whole domain. They are tangential to a triple system of mutually orthogonal curves.

If the triple labels  $[x, y, z]$  of colours are represented graphically by the points of a Euclidean space, with  $x, y, z$ , say, as their Cartesian co-ordinates, then the equation

$$c_{ik}x_ix_k = \text{const.}$$

will represent, on that graph or space model, an *ellipsoid* whose principal axes are those remarkable directions, the principal axes of the tensor  $c_{ik}$ . Referred to these axes, the equation of the ellipsoid is

$$c_{11}'x'^2 + c_{22}'y'^2 + c_{33}'z'^2 = \text{constant.}$$

So much for the salient characteristics of the normal trichromatic frequency distribution law and the methodical way of determining its most probable coefficients from any actual set of free colour matchings.

Our next task is the evaluation of the so-called *standard deviation* (5) of the matches  $(x, y, z)$  from the colour C to be matched, by means of the same normal frequency law.

In the space model just mentioned introduce polar co-ordinates  $r, \theta, \phi$  through

$$x, y, z = r \cos \theta, r \sin \theta (\cos \phi, \sin \phi).$$

Then the frequency function (15) will become

$$f = A e^{-F^2(\theta, \phi)r^2}, \quad \dots \dots \dots (17)$$

where  $A = \sqrt{D/\pi^3}$  and

$$F^2(\theta, \phi) = c_{11} \cos^2 \theta + \sin^2 \theta (c_{22} \cos^2 \phi + c_{33} \sin^2 \phi) + c_{12} \sin 2\theta \cdot \cos \phi + \dots, \quad \dots \dots (17a)$$

a function of the angles  $\theta, \phi$  only, which define a "direction" issuing from C. The coefficients  $c_{11}$ , etc., are the same for all directions, but vary from place to place, *i. e.*, from centre to centre.

The probability corresponding to the frequency (17) is, *per unit solid angle* with apex at C,

$$r^2 f = A r^2 e^{-F^2 \cdot r^2}.$$

Hence, the probable value of the squared charted distance  $r^2$ , within an elementary solid angle around the direction  $\theta, \phi$ ,

$$[r^2]_{\theta, \phi} = A \int_0^\infty r^4 e^{-F^2 r^2} dr \cdot \sin \theta \cdot d\theta \cdot d\phi,$$

which amounts to

$$[r^2]_{\theta, \phi} = \frac{3\sqrt{\pi}A}{8F^5} \sin \theta \cdot d\theta \cdot d\phi.$$

Similarly, we find, for the probable value of  $r$  itself,

$$[r]_{\theta, \phi} = \frac{A}{2F^4} \sin \theta \cdot d\theta \cdot d\phi.$$

These results may be written briefly, *per unit solid angle*,

$$[r^2] = \frac{3A\sqrt{\pi}}{8F^5}, \quad [r] = \frac{A}{2F^4}.$$

In much the same way the probable values of any other expressions, made up of the stimuli-co-ordinates, could be determined. Of especial importance, however, is the probable or mean *squared*  $r$ . The standard deviation  $\sigma$  of the normal distribution, defined by  $\sigma^2=[r^2]$  is (per unit solid angle),

$$\sigma = \sigma_{\theta, \phi} = \sqrt{\frac{3}{8}} \pi^{1/4} \sqrt{\frac{A}{F^5}},$$

or, since  $A=\sqrt{D/\pi^3}$ ,

$$\sigma = \sqrt{\frac{3}{8\pi}} \cdot \frac{D^{1/4}}{F^{5/2}}, \quad \dots \dots \dots (18)$$

where  $D=|c_{ik}|$  and  $F=F(\theta, \phi)$  is as in (17 a). This being the standard deviation of colour matches in a direction  $(\theta, \phi)$ , and the opposite direction being  $\theta^*=\pi-\theta$ ,  $\phi^*=\phi+\pi$ , we have, by the structure of  $F_1$ ,

$$\sigma^*=\sigma.$$

In fine, the standard deviation in two opposite directions is the same.

The *standard surface* (briefly to be called so) surrounding a central colour C, that is, the locus of points (colours) whose charted distance from C is  $\sigma=\sqrt{\xi^2+\eta^2+\zeta^2}$ , say, is expressed by (18). If  $\xi, \eta, \zeta$  be the *rectangular* components of  $\sigma$  along the *principal axes* of  $c_{ik}$ , the equation of this surface is

$$c_1\xi^2+c_2\eta^2+c_3\zeta^2=\frac{3c_1c_2c_3}{8\sqrt{\pi}F}, \quad \dots \dots \dots (19)$$

where

$$F^2=c_1\cos^2\theta+\sin^2\theta(c_2\cos^2\phi+c_3\sin^2\phi)=\frac{1}{\sigma^2}(c_1\xi^2+c_2\eta^2+c_3\zeta^2),$$

where  $c_1=c_{11}$ , etc. are the principal values of the tensor  $c_{ik}$ . Since, for any pair of opposite directions,  $F^*=F$ , the standard surface (19) is, first of all, a *central* surface, C being its centre. It is, however, not an ellipsoid†. The original equation (18) of the standard surface may be

written explicitly :  $\sigma^2=\frac{1}{4}\left(\frac{D}{\pi^3}\right)^{1/2}F^5,$

or 
$$\frac{1}{\sigma^2}=4\sqrt{\frac{\pi^3}{D}}\{c_1\cos^2\theta+\sin^2\theta(c_2\cos^2\phi+c_3\sin^2\phi)\}^{5/2}. \quad \dots \quad (20)$$

† As one might expect by analogy with the ellipses belonging to a two-dimensional *equiluminous* section of the colour domain,

This gives the radius vector  $\sigma$  of any point Q in terms of the direction  $\theta, \phi$  of CQ. With the axes adopted here, the constant factor of the trigonometrical expression is

$$\frac{4\sqrt{\pi^3}}{c_1 c_2 c_3}.$$

The principal axes of the tensor  $c_{ik}$  are

$$\theta = 0 \text{ (and } \theta^* = \pi); \quad \theta = \frac{\pi}{2}, \phi = 0; \quad \theta = \frac{\pi}{2}, \phi = \frac{\pi}{2},$$

and the standard deviations along these axes,

$$\sigma_1 = 2\pi^{3/4} \frac{\sqrt{c_2 c_3}}{c_1^2}, \text{ etc.,}$$

by cyclic permutation of the subscripts 1, 2, 3.

On the other hand, the necessary conditions,

$$\frac{\partial \sigma}{\partial \theta} = \frac{\partial \sigma}{\partial \phi} = 0,$$

for a maximum or minimum of  $\sigma$  are

$$\sin^2 \theta [c_2 \cos^2 \phi + c_3 \sin^2 \phi - c_1^{3/4}] = 0,$$

and

$$\sin^2 \theta \cdot \sin^2 \phi = 0.$$

The first equation has the roots

$$\theta_1 = 0 \text{ (any } \phi), \quad \theta_2 = \frac{\pi}{2} \text{ (any } \phi), \quad \text{and} \quad \cos^2 \phi_3 = \frac{c_3 - c_1}{c_3 - c_2} \text{ (any } \theta),$$

and the second equation,

$$\theta_1 = 0 \text{ (any } \phi), \quad \phi_2 = 0 \text{ (any } \theta), \quad \phi_3 = \frac{\pi}{2} \text{ (any } \theta).$$

Ultimately, therefore, we have, first of all, the direction  $\theta_1 = 0$  coinciding with the first principal axis of  $c_{ik}$  and the direction

$$\theta_2 = \frac{\pi}{2}, \quad \phi_2 = 0$$

coinciding with the second axis of this tensor, which leaves only

$$\phi_3 = \frac{\pi}{2} \quad \text{and} \quad \cos^2 \phi_3 = \frac{c_3 - c_1}{c_3 - c_2} \text{ (any } \theta). \quad . . . . \text{ (N)}$$

Now, if  $c_3 \neq c_1$ , these two conditions are incompatible and we are left only with extremal values of  $\sigma$  along two perpendicular axes of the tensor,

$$\theta = 0 \text{ and } \theta = \frac{\pi}{2}, \quad \phi = 0,$$

namely, as we have just seen,

$$\sigma_1 = \frac{2\pi^{3/4} \sqrt{c_2 c_3}}{c_1} \quad \text{and} \quad \sigma_2 = \frac{2\pi^{3/4} \sqrt{c_3 c_1}}{c_2},$$

which give the ratio  $\sigma_1^2 : \sigma_2^2 = (c_2 : c_1)^3$ . If, say,  $c_2 > c_1$ , then  $\sigma_1 > \sigma_2$ , the first value is a maximum and the second, a minimum of  $\sigma$ . The section of the standard surface laid through these two axes has the comparatively simple equation

$$\frac{1}{\sigma^2} = 4 \sqrt{\frac{\pi^3}{D}} \{c_1 \cos^2 \theta + c_2 \sin^2 \theta\}^{5/2}. \quad (20.1)$$

Returning to the solutions (N), we see at once that  $\sigma$  can have a third extremal value only provided that  $\cos \phi_3 = 0$  and, therefore,

$$c_3 = c_1.$$

(Two principal values of our tensor equal.) In this particular case the surface (20) becomes

$$\frac{1}{\sigma^2} = 4 \sqrt{\frac{\pi^3}{D}} \{c_1 (\cos^2 \theta + \sin^2 \theta \cdot \sin^2 \phi) + c_2 \sin^2 \theta \cdot \cos^2 \phi\}^{5/2}. \quad (20.2)$$

Since  $\sigma^2 = \xi^2 + \eta^2 + \zeta^2$ ,  $\sigma \cos \theta = \xi$ , etc., the equation of the standard surface is always of the *fourteenth* degree in  $\xi, \eta, \zeta$  or of the *seventh* degree in  $\xi^2, \eta^2, \zeta^2$ , and is thus far from being a quadric, although it may sometimes resemble an ellipsoid. In general, the surface is not even everywhere convex. It may very well have concave portions.

In the subcase  $c_2 = c_3 \neq c_1$ , when the solution (N) is manifestly not available, we have, by (20),

$$\frac{1}{\sigma^2} = 4 \sqrt{\frac{\pi^3}{D}} \{c_1 \cos^2 \theta + c_2 \sin^2 \theta\}^{5/2}, \quad (20.3)$$

a *surface of revolution*, with the extremal  $\sigma$ -values,

$$\sigma_1 = \frac{1}{2c_1} \left( \frac{D}{\pi^3} \right)^{1/4}, \quad \sigma_2 = \frac{1}{2c_2} \left( \frac{D}{\pi^3} \right)^{1/4}$$

corresponding to  $\theta = 0$  and  $\theta = \frac{\pi}{2}$ .

The meridian curve (20.3) of this axially symmetrical surface is still of the *seventh* degree in  $\xi^2, \eta^2$ .

An example, corresponding to  $c_1 = 4$ ,  $c_2 = 1$ , is shown in fig. 1. The meridian curve of the standard surface resembles rather a lemniscate than an ellipse. The equatorial plane is a plane of symmetry of the surface.

If also  $c_1 = c_2$ , and thus  $c_1 = c_2 = c_3$ , equal to  $c$  say, then our surface of revolution becomes a sphere,

$$\sigma^2 = \frac{1}{4} \sqrt{\frac{D}{\pi^3}} \cdot \frac{1}{c^5} = \frac{1}{4\pi^{3/2} c^{7/2}}.$$

Its radius  $\sigma$  is inversely proportional to  $c^{7/4}$ .



The frequency distribution of matches, for unit solid angle, around the central colour C is, by what precedes,

$$f(\theta, \phi) = \frac{c_1 c_2 c_3}{4\pi F^3}, \quad \dots \dots \dots (21)$$

where F is the function of  $\theta, \phi$  just defined. This holds generally for any tensor  $c_{ik}$ .

The distribution of colour matches along a radial direction ( $\theta, \phi$  constant) issuing from  $c$ , briefly, their *radial distribution*, is expressed by

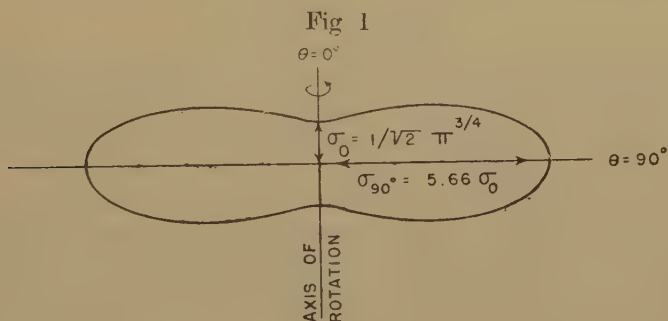
$$f_{\theta, \phi \text{ const.}}(r) = \frac{c_1 c_2 c_3}{\pi^{3/2}} e^{-F^2 \cdot r^2} \dots \dots \dots (22)$$

The probability of a match falling between  $r$  and  $r+dr$  is

$$r^2 f(r) dr \cdot \sin \theta \cdot d\theta \cdot d\phi = \text{const.} \cdot r^2 e^{-F^2 r^2} dr, \quad \dots \dots (22a)$$

one way, and twice as much if taken in the two opposite directions.

*Metrics.*—The second-rank tensor  $c_{ik}$ , introduced through the assumed invariance of the probability of colour matchings, enables us to *metricize* the whole colour domain.



If  $P(x, y, z)$  and  $Q(x+dx, \text{etc.})$  be two contiguous colours, then, as stated before,  $\overrightarrow{PQ}$  or  $dx, dy, dz$  is a general (contravariant) tensor of rank one or vector, but it has no invariant of its own,  $\overline{PQ} = \sqrt{dx^2 + dy^2 + dz^2}$  being merely the casual, charted, distance of the two colours. Our tensor  $c_{ik}$ , however, will at once endow it with an invariant, viz.,  $c_{11}dx^2 + 2c_{12}dxdy + \dots + c_{33}dz^2$ . In conformity with formula (7), we may define the *true distance*  $ds$  between the two colours  $(x, y, z)$  and  $(x+dx, \text{etc.})$  by

$$ds^2 = \frac{1}{m} \{c_{11}dx^2 + 2c_{12}dxdy + \dots + c_{33}dz^2\}, \quad \dots \dots (23)$$

or, concisely,

$$ds^2 = \frac{1}{m} c_{ik} dx^i dx^k,$$

where  $m$  is some conventional pure number.

The frequency or probability of matching per unit volume of the domain is, by (11), with  $f(x)$  written for  $f(x_1, x_2, x_3)$ ,

$$f(x) = f(0) e^{-c_{\iota\kappa} dx^{\iota} dx^{\kappa}} = f(0) e^{-m ds^2}.$$

Hence, we may represent the true distance of two colours  $P(x)$  and  $Q(x+dx)$  by

$$ds^2 = \frac{1}{m} \log \frac{f(0)}{f(x)}. \quad (24)$$

The "true" or invariant distance between colours in the three-dimensional domain is thus expressed by a simple function of matching probabilities, as in a previous paper\* on the two-fold of equiluminous colours, a section ( $x+y+z=\text{constant}$ ) of the full domain. In that chance-section† the distance element  $ds$  happened also to be  $\overline{PQ}/\sigma$ , that is, equal to the charted distance with the standard deviation  $\sigma = \sigma_{\theta, \phi}$  as unit. In the complete, three-dimensional domain the relation between the true distance ( $ds$ ) and the charted one ( $\overline{PQ}$ ) and the standard deviation ( $\sigma$ ) in the direction  $\overrightarrow{PQ}$  is less simple, to wit, by the definition (23) and our  $\sigma$ -formula,

$$ds = D^{1/10} \frac{\overline{PQ}}{\sigma^{2/5}}, \quad (25)$$

where  $D$  is the determinant of  $c_{\iota\kappa}$ , now the metrical tensor of the colour domain. Apart from the factor  $D^{1/10}$  (independent of direction), the previous standard deviation in an equiluminous section is here replaced by its fractional power  $\sigma^{2/5}$ .

The most significant interpretation of "true distance" between colours in terms of experimental, observable peculiarities is embodied in (24), where  $m$  is a mere number, to be chosen at our pleasure. If we take, simply,  $m=1$ , then the squared distance element is

$$ds^2 = \log \frac{f(0)}{f(x)} = \log \frac{P(0)}{P(x)}.$$

These are, essentially, all the relevant results of the proposed tensorial treatment of the colour three-fold.

In closing this subject, for the present, we should like to urge extensive colour matchings in the *complete* three-dimensional domain. It seems that these and these only are apt to disclose the characteristics of the fascinating domain of our colour sensations. Neither a single (equiluminous) section  $x+y+z=h=\text{const.}$ , nor even a plurality of such sections,  $x+y+z=h_1$ ,  $x+y+z=h_2$ , and so on, can ever lead to an adequate knowledge of the domain as a whole, of its intrinsic, characteristic properties. Free colour matchings in the complete three-fold, moreover, do not seem to offer any greater difficulties than those

\* L. Silberstein and D. L. MacAdam, J. Opt. Soc. Amer. vol. xxxv. pp. 32-39 (January, 1945), Kodak Comm. No. 995.

† "Chance," since there is nothing intrinsic about  $x+y+z$ .

confined deliberately to equiluminous or other two-dimensional sections of the domain.

*Supplement: On Free and Guided Colour Matchings.*

Let us, first, sum up the main results just found for the complete domain.

The probability that a free match of a colour C shall fall within a solid angle  $\sin \theta \cdot d\theta \cdot d\phi$ , with apex at C and at a charted distance  $r$  to  $r+dr$  from C is  $f r^2 \sin \theta \cdot d\theta \cdot d\phi$ , where

$$f = \sqrt{\frac{D}{\pi^3}} e^{-F^2(\theta, \phi) r^2}.$$

If the  $x$ -,  $y$ -,  $z$ -axes coincide, locally, with the principal axes of the tensor  $c_{\kappa\kappa}$ , then

$$F^2 = c_1 \cos^2 \theta + \sin^2 \theta (c_2 \cos^2 \phi + c_3 \sin^2 \phi),$$

where  $c_\kappa$  stands for  $c_{\kappa\kappa}$  ( $\kappa=1, 2, 3$ ).

The probable value of  $r^2$  along a fixed direction  $(\theta, \phi)$  is

$$[r^2] = \frac{3}{8} \frac{\sqrt{D}}{\pi} \cdot \frac{1}{F^5} (\sin \theta \cdot d\theta \cdot d\phi),$$

or, per unit solid angle,

$$\sigma^2 = \frac{[r^2]}{\sin \theta \cdot d\theta \cdot d\phi} = \frac{3\sqrt{D}}{8\pi F^5}. \quad (A)$$

The corresponding surface, with  $\sigma$  as radius vector, is of the seventh degree in the squared co-ordinates  $U, V, W = \sigma(\cos \theta, \sin \theta \cdot \cos \phi, \sin \theta \cdot \sin \phi)$ .

Formula (A) holds, of course, for the distribution of *free* (not guided) matches, giving the squared standard deviation of such matches, that is, within a cone of constant solid angle.

It is of interest to determine an analogous magnitude, not for a cone, but for a *cylinder*, of constant, infinitesimal transversal section,  $\alpha$  say, whose axis drawn through C has any given direction  $(\theta, \phi)$ . This will cover the case of *guided matchings*, viz., constrained to remain within the cylinder specified. The probability that such a guided match shall fall within a charted distance  $r$  to  $r+dr$  from C is for  $f(r)dr$ ,

$$f(r) = c e^{-F^2(\theta, \phi) \cdot r^2};$$

where  $F$  is the function of  $\theta, \phi$  defined before, and  $c$  a factor independent of  $r$  (and of  $\theta, \phi$ )\*. Now, the probability that a guided match shall fall within the said cylinder at *any* distance whatever from C, either along

---

\* If  $\alpha$  is the area of the section of the cylinder and, therefore,  $\alpha dr$  the volume element, we should have, for *free* matches,  $c = \alpha \sqrt{\frac{D}{\pi^3}}$ .

$\theta$ ,  $\phi$  or in the opposite direction, is obviously equal to 1, since all matches are forced to remain within that cylinder. Thus,

$$1 = 2 \int_0^\infty f(r) dr = 2c \int_0^\infty e^{-F^2 r^2} dr,$$

or 
$$\frac{2c}{F} \int_0^\infty e^{-u^2} du = 1,$$

whence 
$$c = \frac{F}{\sqrt{\pi}}.$$

Thus, the frequency of the guided matches becomes

$$f(r) = \frac{F}{\sqrt{\pi}} e^{-F^2 r^2}.$$

and the probable value of  $r^2$  or the squared standard deviation of *guided matches*,

$$\sigma_g^2 = \frac{2\pi}{\sqrt{\pi}} \int_0^\infty r^2 e^{-F^2 r^2} dr = \frac{1}{2F^2},$$

or 
$$\frac{1}{\sigma_g^2} = 2F^2 = c_1 \cos^2 \theta + \sin^2 \theta (c_2 \cos^2 \phi + c_3 \sin^2 \phi). \quad \dots \quad (B)$$

In fine, while for free matches, *i.e.*, within a cone of constant solid angle,  $\sigma^2 \sim 1/F^2$ , for guided matches constrained to a cylinder of constant section ( $\alpha$ ),

$$\sigma_g^2 \sim \frac{1}{F^2},$$

and the corresponding surface (locus of endpoints of the radius vector  $\sigma_g$ ) is an *ellipsoid*. If we write, as before,  $U = \sigma_g \cos \theta$ , etc. and notice that  $F^2 \sigma_g^2 = \frac{1}{2}$ , the equation of the ellipsoid will be

$$c_1 U^2 + c_2 V^2 + c_3 W^2 = \frac{1}{2},$$

where  $c_\kappa = c_{\kappa\kappa}$  are the principal values of the metrical tensor. Its principal axes are the principal axes of the ellipsoidal surface surrounding the colour  $C$  to be matched, the surface being (as before) the locus of (the endpoints of) standard deviations of the guided matchings of that colour. With this particular definition of the standard deviation ( $\sigma_g$ ), the true or the invariant distance of two colours,  $P(x)$  and  $Q(x+dx)$ , is proportional to  $\overline{PQ} \sigma_g$ , where  $\overline{PQ} = \sqrt{dx^2 + dy^2 + dz^2}$  is their charted distance, in perfect harmony with the definition of the distance  $ds$  given (*loc. cit.*) for the two-fold of equiluminous colours.

Rochester, New York.  
January 20, 1945.

---

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



FIG. 1.



Photograph of great sunspot, taken at Royal Observatory, Greenwich, 5th February, 1946.



XIX. *A Problem on the Summation of Simple Harmonic Functions of the same Amplitude and Frequency but of Random Phase.*

By F. HORNER, M.Sc.

(Radio Division, National Physical Laboratory).

[Received April 28, 1945.]

ABSTRACT.

The following problem has arisen in the theoretical study of site errors in direction finding.

"A number of simple harmonic functions of the same amplitude and frequency, but having phase angles which may have any equally probable value between zero and  $2\pi$ , are added together. That component of the resultant simple harmonic function is found which has an arbitrarily chosen phase. What is the probability that the amplitude of this component will lie between any assigned limits?"

The problem is similar to that of the "Random Walk," to which solutions have been given by previous workers, and a brief resumé is given of the Random Walk problem and its solution. The solution of the present problem is discussed in terms of the probability function  $P_n(s)$ , where  $P_n(s)ds$  is the probability that the amplitude of the chosen component will lie between  $s$  and  $s+ds$  for a given number  $n$  of harmonic functions. It is shown that  $P_n(s)$  is not in general expressible in terms of well-known functions, and it must then be calculated by graphical means or by the use of infinite series, which involve very little labour if  $n$  is greater than six.

Curves of  $P_n(s)$  are given as functions of  $s$  for  $n=1, 2, 3$  and  $7$ . The curves for  $n=1, 2$  and  $3$  have infinities or discontinuities in slope. An approximate solution is also given, which is applicable when  $n$  is large. This is of normal or Gaussian form, and the true probability distributions tend progressively towards the normal form as  $n$  is raised. The curve for  $n=7$  is very close to the normal form, and it seems likely that the lowest value of  $n$ , for which the normal curve gives a reasonably good representation of the true distribution, is five. The root mean square values of  $s$  are the same for the true and normal distributions for all values of  $n$ , and the arithmetic mean and probable values are the same for both distributions within a few per cent. for values of  $n$  greater than one.

Brief mention is made of the application of this theoretical investigation to site error problems in direction-finding. The results are of use in determining the probable bearing error when a number of interfering rays of equal amplitude, but completely random phase, are incident to a direction finder from a narrow range of azimuth. More generally the

theory is applicable when the interfering rays arrive from any azimuths whatever, so long as the amplitude of the pick-up on the direction finder is the same from each of these rays. In this case it will, in general, be necessary to assume that errors are small, and that the pick-up from each interfering ray is sensibly independent of small displacements of the aerial system from its "on-bearing" position.

### 1. *Introduction.*

THE following problem has arisen in the theoretical study of site errors in direction finding.

"A number of simple harmonic functions of the same amplitude and frequency, but having phase angles which may have any equally probable value between zero and  $2\pi$ , are added together. That component of the resultant simple harmonic function is found which has an arbitrarily chosen phase. What is the probability that the amplitude of this component will lie between any assigned limits?"

It may happen, for instance, that a group of similar obstacles on or near to a direction-finding site causes bearing errors. If the obstacles are close together, compared with their distance from the direction-finder, but not compared with the wave-length, it may be possible to regard the waves reflected by the individual obstacles as all having the same amplitude at the direction-finder but phase angles depending on the exact location of the obstacles. If the amplitude of the reflected wave due to any one obstacle is known, but the relative phase angles of the various interfering waves are unknown or are changing with time (*e.g.* due to balloon barrage cables moving in the wind), the solution to the problem stated above is needed to calculate the probable magnitude of the resultant error.

The harmonic functions may be represented by  $n$  coplanar vectors of equal length  $l$ , but arbitrary directions. The probability distribution of the magnitude of the projection of the resultant vector on an arbitrarily chosen straight line is required\*. In the site error problem the direction of this line represents the phase at the direction-finder of the wave received direct from the transmitter†.

The problem is fundamentally a modification of that which was called by Professor Pearson "The Problem of the Random Walk" <sup>(1)</sup>, the solution to which was given by Pearson, Kluyver and Rayleigh <sup>(2, 3, 4, 5)</sup>. It will be instructive to compare the two problems and their solutions.

---

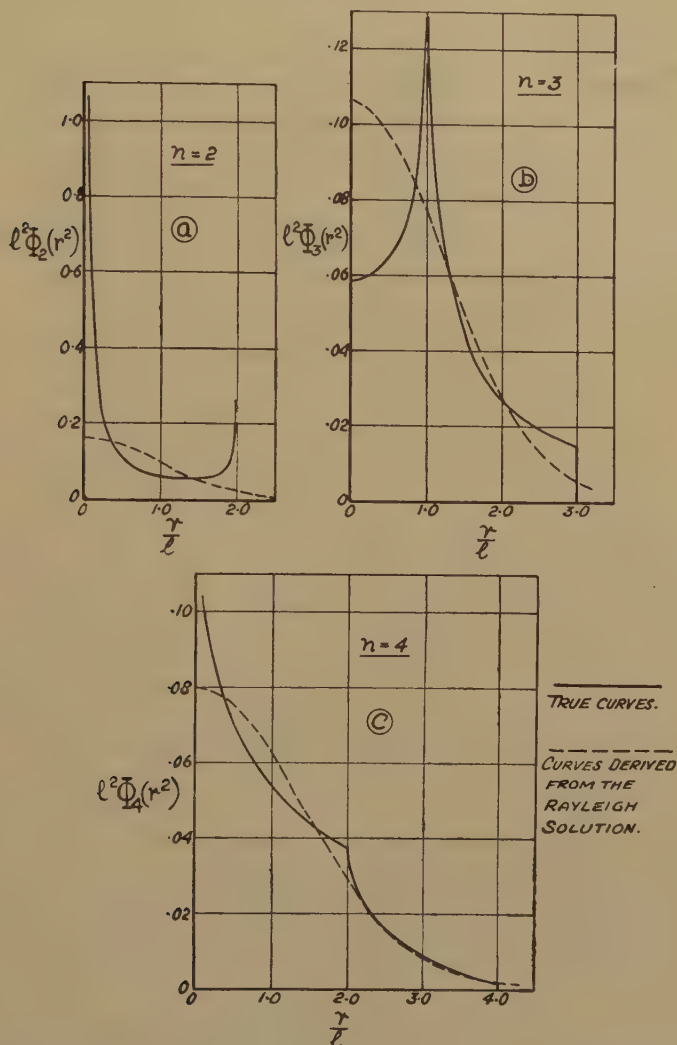
\* The probability distribution of the instantaneous value of the projection is independent of the frequencies of the individual vectors (which may be all different) provided the initial phases are random. The direction finding problem, however, specifically relates to a number of oscillations of the same frequency in which case the projection represents the amplitude of that component of the resultant simple harmonic function with the specified phase.

† It is assumed that the phase of the pick-up on the direction finder is dependent only on the phase of the incident wave and not on its direction of arrival.



The bulk of Pearson's work on the Random Walk problem was carried out in connection with the migration of species, and was published in a rather inaccessible journal <sup>(2)</sup>. A brief résumé will therefore be given of this problem and of its solution.

Fig. 1 (a-c).



Dispersal curves (Pearson).

## 2. Résumé of Work on Pearson's Random Walk Problem.

Pearson's problem was stated in the following terms <sup>(1)</sup> :—

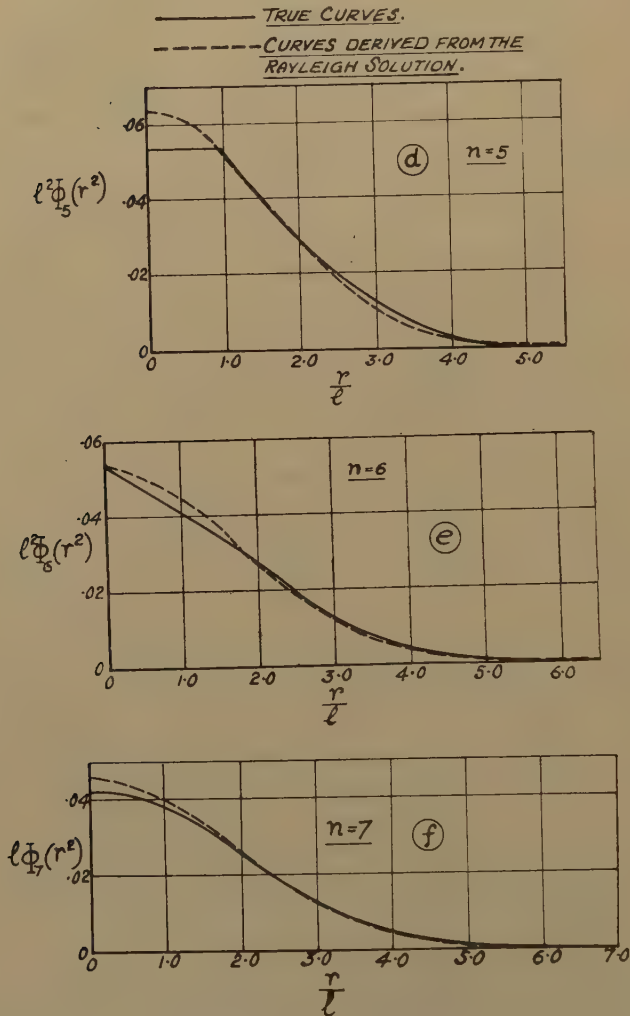
"A man starts from a point O and walks  $l$  yards in a straight line; he then turns through any angle whatever and walks another  $l$  yards in a second straight line. He repeats this process  $n$  times. I require

the probability that after these  $n$  stretches he is at a distance between  $r$  and  $r+dr$  from his starting point O."

The solution of this problem for large values of  $n$  was shown by Rayleigh <sup>(3)</sup> to be closely represented by

$$\Phi_{\infty}(r^2) = \frac{1}{\pi n l^2} e^{-r^2/nl^2}, \quad \dots \dots \dots (1)$$

Fig. 1 (d-f).



Dispersal curves (Pearson).

where  $\Phi_{\infty}(r^2)$  is the theoretical frequency of the man's position on unit area at a distance  $r$  from his starting point, and  $e$  is the base of natural logarithms. This result is correct only when  $n$  is infinite (hence the subscript  $\infty$ ), but it gives a close approximation to the true distribution

for large, finite values of  $n$  over the range  $r=0$  to  $r=nl$ . The frequency must necessarily be zero for values of  $r$  greater than  $nl$ .

Pearson set out to determine how large  $n$  must be before the Rayleigh solution gives a close approximation to the true distribution. The case  $n=1$  is trivial, since the man's distance from the origin must be  $l$  after one stretch. He found that the case  $n=2$  could be solved in terms of elementary functions, and the case  $n=3$  by the use of elliptic integrals. Solutions for larger finite values of  $n$  cannot be expressed in terms of well-known functions, and Pearson developed infinite series from which the frequency distributions could be calculated. He also developed a graphical method for obtaining  $\Phi_{n+1}(r^2)$  from  $\Phi_n(r^2)$ . Both these methods entail considerable labour, but he derived the curves for  $n=4, 5, 6$  and  $7$  by the graphical method and for  $n=6$  and  $7$  also by series. His results are reproduced in the curves of fig. 1, which show  $l^2 \Phi_n(r^2)$  plotted against  $r/l$ . The corresponding curves derived from equation (1) are also shown, and the following characteristics are noteworthy:—

- (i.) For very small values of  $n$  the Rayleigh solution is widely divergent from the true distribution, and the frequency curves have infinities or discontinuities in slope.
- (ii.) The frequency is always finite for  $n \geq 5$ .
- (iii.) The true distribution tends progressively towards the Rayleigh form as  $n$  is raised.
- (iv.) For  $n=7$  there are no apparent discontinuities, and the curve is similar in form to that given by the Rayleigh solution, although there is a small discrepancy between the true curve and that deduced from equation (1). Pearson concluded that for all practical purposes the Rayleigh solution is a true representation of the facts for  $n \geq 10$ .

The solution of the problem by means of series becomes less and less tedious as  $n$  is raised, and is reasonably simple for  $n \geq 6$  with the aid of certain functions which Pearson has tabulated.

It is to be expected that similar conclusions are valid for the solution of the problem stated in Section 1, and the following sections are devoted to a determination of a partial solution. Comparison is made with the solution of Pearson's problem, which for ease of reference is referred to as problem A, that stated in Section 1 being termed problem B.

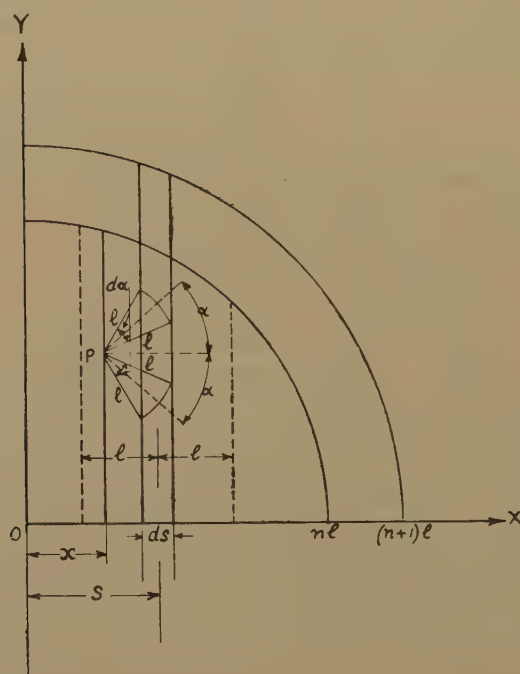
### 3. The Solution of Problem B.

A comparison of problem B with problem A, as stated in Section 2, shows that they are the same except that in problem B we require to find the man's probable distance, not from his starting point, but from an arbitrary straight line through it. It follows that if we know the probability distribution of his position relative to his starting point, from Pearson's curves, then the solution to problem B can be determined by an integration process. It can be inferred that the solutions to

problem B will have infinities or discontinuities in slope when  $n$  is small, and that a similar solution to Rayleigh's solution of problem A can be obtained when  $n$  is very large. Owing to the complexity of the functions for small finite values of  $n$  greater than 2 the derivation of solutions of problem B from Pearson's curves is somewhat laborious, especially as it is necessary to measure areas under curves of functions which become infinite in the range of integration.

The solution of problem B for small finite values of  $n$  has therefore been approached from first principles. A general functional relationship between the probability functions of orders  $n$  and  $(n+1)$  has been found

Fig. 2.



and solutions determined for  $n=1, 2$  and  $3$ . In addition a solution for  $n=7$  has been derived from the series developed by Pearson.

#### 4. *The Solution of Problem B for Large Values of $n$ .*

This can be derived most simply from Rayleigh's solution of problem A quoted in equation (1). Using the vector representation given in Section 1, this equation gives the frequency of the terminal point of the resultant vector on unit area at distance  $r$  from the origin. It is a matter of simple integration to find the frequency of the terminal point on a strip of width  $ds$  parallel to and distance  $s$  from any arbitrary straight line through the origin. (See fig. 2, in which the Y-axis is the arbitrary straight line.)

Let  $P_n(s)ds$  be the probability that the point lies in this strip for a given



value of  $n$ . Then, since the length of the strip inside a circle of radius  $nl$  centred on the origin is  $2\sqrt{n^2l^2 - s^2}$  (only half this strip is shown),

$$P_n(s) = 2 \int_0^{\sqrt{n^2l^2 - s^2}} \Phi_n(r^2) dy.$$

If  $n$  is large we can use the Rayleigh solution for  $\Phi_n(r^2)$  and

$$P_n(s) = \frac{2}{\pi n l^2} e^{-s^2/nl^2} \int_0^{\sqrt{n^2l^2 - s^2}} e^{-y^2/nl^2} dy,$$

or, if  $z^2 = y^2/nl^2$ ,

$$P_n(s) = \frac{2}{\pi l \sqrt{n}} e^{-s^2/nl^2} \int_0^{\sqrt{n - s^2/nl^2}} e^{-z^2} dz. \quad . \quad . \quad . \quad (2)$$

If we let  $n$  tend to infinity the upper limit of integration tends to infinity, and we have for very large values of  $n$

$$P_\infty(s) = \frac{1}{l \sqrt{(\pi n)}} e^{-s^2/nl^2}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This probability distribution has the normal form and, like equation (1), is correct only if  $n$  is infinite. If  $n$  is large it gives a close approximation to the true distribution for values of  $s$  between  $-nl$  and  $+nl$ ; outside these limits the probability function is necessarily zero.

We shall be concerned with finding the smallest value of  $n$  for which the normal law gives a close approximation to the true curve, and comparison will be made between the arithmetic mean, root mean square and probable values of  $s$  as derived from the true and normal distributions. The probable value is defined as that value which is just as likely as not to be exceeded by a positive value of  $s$ .

Since the true probability function is zero when  $s$  lies outside the limits  $\pm nl$ , it might seem appropriate to compare the true distribution with that part of the normal distribution lying between these limits. Practical considerations, however, make it desirable to compare the true distribution with the whole of the normal curve in respect of the above features. The normal curve then has the correct area and standard deviation (R.M.S. value), and the arithmetic mean and probable values of  $s$  can be expressed in simple terms. This procedure will thus indicate when it is possible to use simple expressions to define the properties of the true curves to a close approximation. Moreover, even when the value of  $n$  is only moderately large, almost the whole of the area under the normal curve lies between the limits  $s = \pm nl$ , and for practical purposes the effect of including that part of the curve outside the limits is negligible.

In the remainder of this paper the R.M.S., arithmetic mean and probable values of  $s$  will therefore be derived from the whole of the normal curve for which

$$\text{R.M.S. value} = l \sqrt{\left(\frac{n}{2}\right)},$$

$$\text{Arithmetic mean value} = 0.5643 l \sqrt{n}.$$

$$\text{Probable value} = 0.4769 l \sqrt{n}.$$

5. *Root Mean Square Value of the Displacement s.*

The R.M.S. value of the displacement  $s$  can be derived simply for any value of  $n$ , and is a useful quantity in the study of direction finding errors. If  $\phi_1, \phi_2, \dots, \phi_n$  are the arbitrary phase angles of the  $n$  vectors, referred to the direction in which  $s$  is measured, then

$$s = l(\cos \phi_1 + \cos \phi_2 + \dots + \cos \phi_n). \quad (4)$$

In  $N$  trials there are

$$\frac{N}{(2\pi)^n} d\phi_1 \cdot d\phi_2 \dots d\phi_n$$

resultants having phase angles in the small limits  $d\phi_1, d\phi_2, \dots, d\phi_n$  at  $\phi_1, \phi_2, \dots, \phi_n$ , and these have a displacement  $s$  given by equation (4). The mean square value of  $s$  is therefore

$$\frac{l^2}{(2\pi)^n} \int_0^{2\pi} d\phi_1 \int_0^{2\pi} d\phi_2 \dots \int_0^{2\pi} [\cos \phi_1 + \cos \phi_2 + \dots + \cos \phi_n]^2 d\phi_n, \quad (5)$$

which by successive integration may be reduced to  $\frac{1}{2}nl^2$ . Thus, for all values of  $n$ ,

$$\text{R.M.S. value of } s = l \sqrt{\left(\frac{n}{2}\right)}. \quad (6)$$

6. *General Functional Relationship for Problem B.*

A relationship will be developed between  $P_{n+1}(s)$  and  $P_n(s)$ . Fig. 2 shows a plane on which a large number ( $N$ ) of points may be plotted, each representing the terminal point of the resultant of  $n$  vectors each of length  $l$ . The points may be regarded as having started at the origin  $O$  and having undergone  $n$  displacements of equal length  $l$  but arbitrary directions. If  $N$  is sufficiently large the distribution conforms to Pearson's curve of order  $n$ , *i. e.* the density at a point  $P(x, y)$  is  $N\Phi_n(r^2)$ , where  $r = OP$ . The density is zero outside a circle of radius  $nl$ .

Each of the points is now given one more displacement  $l$  to form a new distribution extending up to a circular boundary of radius  $(n+1)l$ . Consider the small strip of width  $ds$  at  $x=s$  shown on the figure, where  $s$  is positive. The points on a small area  $dxdy$  at  $P$  after  $n$  displacements may contribute to the population of the strip after  $(n+1)$  displacements if  $(s-l) < x < (s+l)$ . Since there is equal probability of the displacement of a point in any direction the number of points from the small area at  $P$ , which are in the strip after the  $(n+1)$ th displacement, is

$$N\Phi_n(r^2) dxdy \frac{d\alpha}{\pi},$$

where

$$d\alpha = \frac{ds}{l \sin \alpha} = \frac{ds}{\sqrt{l^2 - (x-s)^2}}, \quad (7)$$

The total number of points on the strip after  $(n+1)$  displacements is therefore

$$2 \int_{s-l}^x \left[ \int_0^{\sqrt{n^2 l^2 - x^2}} N\Phi_n(r^2) \frac{ds}{\pi \sqrt{l^2 - (x-s)^2}} dy \right] dx, \quad (8)$$

where the upper limit  $X$  is determined by the value of  $s$ . Since  $\Phi_n(r^2)$  is zero if  $r > nl$ ,

$$X = s + l \text{ or } nl, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

whichever is smaller.

Now the total number of points on the strip  $ds$  after  $(n+1)$  displacements is  $N P_{n+1}(s) ds$  if  $ds$  is indefinitely small.

Also 
$$2 \int_0^{\sqrt{n^2 l^2 - x^2}} \Phi(r^2) dy = P_n(x). \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Therefore 
$$P_{n+1}(s) = \frac{1}{\pi} \int_{s-l}^X \frac{P_n(x)}{\sqrt{l^2 - (x-s)^2}} dx. \quad . \quad . \quad . \quad . \quad (11)$$

Moreover, since  $s - x = l \cos \alpha$ , (11) is expressible in the alternative form

$$P_{n+1}(s) = \frac{1}{\pi} \int_0^{\cos^{-1} \left( \frac{s-X}{l} \right)} P_n(s - l \cos \alpha) d\alpha. \quad . \quad . \quad (12)$$

Either of these equations may be used to derive the  $(n+1)$ th distribution from the  $n$ th, but in general it is not possible to carry out the integration process except by a graphical method or by the use of infinite series.

Only positive values of  $s$  have been considered since the distributions are symmetrical about the line  $s=0$ .

#### 7. The Solution for $n=1$ .

When  $n=1$ ,  $s = l \cos \phi_1$ , where all values of  $\phi_1$  between 0 and  $2\pi$  are equally probable. In  $N$  trials the number of points lying in a strip of small width  $ds$  at  $s$  is

$$\frac{Nd\phi_1}{\pi} = \frac{Nds}{\pi l \sin \phi_1}.$$

The probability function is therefore

$$P_1(s) = \frac{1}{\pi \sqrt{l^2 - s^2}}. \quad . \quad . \quad . \quad . \quad . \quad (13)$$

In fig. 3,  $lP_1(s)$  is plotted against  $s/l$ . We can also derive the following values :—

$$\text{R.M.S. value of } s = \frac{l}{\sqrt{2}} [\text{from equation (6)}] = 0.707l.$$

$$\text{Arithmetic mean value} = \frac{2l}{\pi} \int_0^{\pi/2} \cos \phi_1 d\phi_1 = \frac{2l}{\pi} = 0.637l.$$

The probable value,  $p$ , is given by

$$\frac{2l}{\pi} \int_0^{\cos^{-1} p} d\phi_1 = \frac{1}{2} \quad \text{or} \quad p = \frac{l}{\sqrt{2}} = 0.707l.$$

#### 8. The Solution for $n=2$ .

From equations (11) and (13)

$$P_2(s) = \frac{1}{\pi^2} \int_{s-l}^l \frac{1}{\sqrt{l^2 - x^2} \sqrt{l^2 - (x-s)^2}} dx. \quad . \quad . \quad . \quad (14)$$

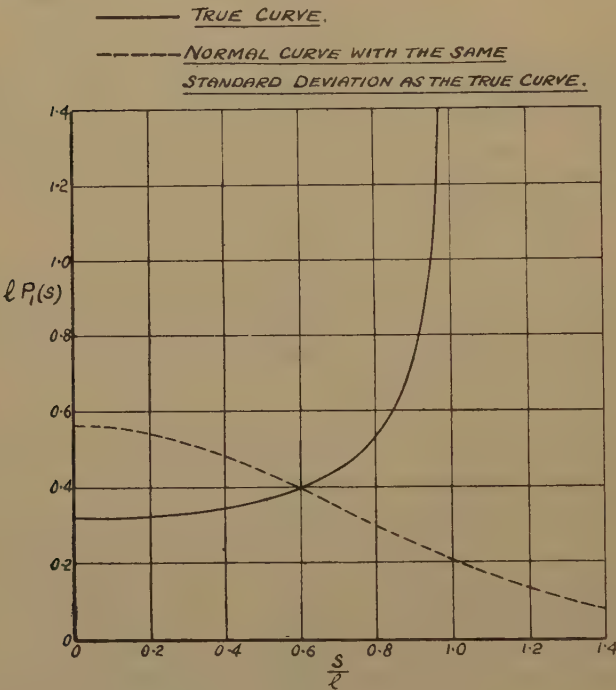
The integral is elliptic and can be reduced to the standard Lagrangian form by appropriate substitutions <sup>(6)</sup>. Then

$$P_2(s) = \frac{1}{\pi^2 l} \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - \sin^2 \theta \sin^2 \phi}} \quad \dots \quad (15)$$

$$= \frac{F(\theta)}{\pi^2 l}, \quad \dots \quad (16)$$

where  $\cos \theta = s/2l$  and  $F(\theta)$  is a complete elliptic integral of standard form tables of which have been published <sup>(7)</sup>.  $lP_2(s)$  is plotted against  $s/l$  in fig. 4. It can be shown that the probability function becomes infinite at

Fig. 3.



Probability distribution for  $n=1$ .

$s=0$  but there are no discontinuities between  $s=0$  and  $s=2l$ , since the upper limit of integration in equation (14) is always less than  $(s+l)$ .

An alternative method of approach to the problem is of interest. Suppose values of the arbitrary phases  $\phi_1, \phi_2$  are plotted along two mutually perpendicular axes, each angle ranging from  $-\pi$  to  $+\pi$ . Then a figure is obtained which consists of a square of which one-quarter is shown in fig. 5. If we plot on this figure a large number of points, each representing the result of a selection of  $\phi_1$  and  $\phi_2$ , there will be a uniform distribution of such points on the figure since all values of  $\phi_1$  and  $\phi_2$  between  $-\pi$  and  $+\pi$  are equally likely to occur. We can also draw lines on the figure



representing values of  $s=l(\cos \phi_1 + \cos \phi_2)$  and two such lines are drawn corresponding to  $s$  and  $s+ds$ . The area enclosed between these two lines and the axes is a measure of the frequency of the result of the trial within these limits, and can be expressed in the same form as equation (14).

Fig. 4.

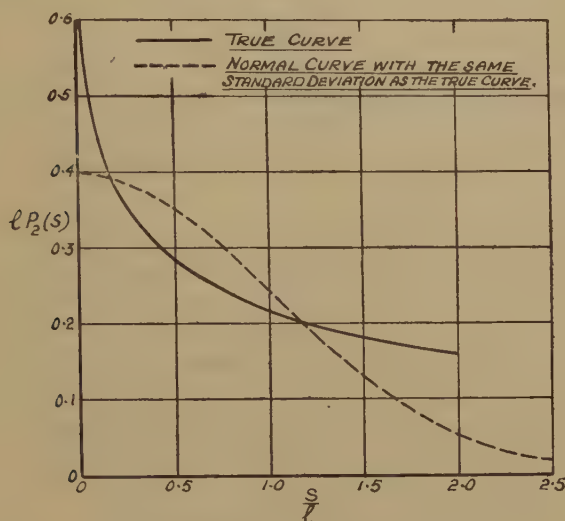
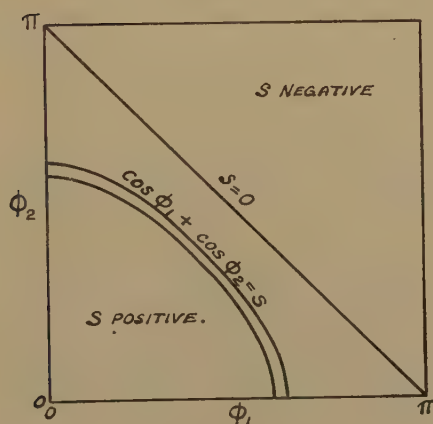
Probability distribution for  $n=2$ .

Fig. 5.



This method of approach leads to a simple expression for the arithmetic mean value of  $s$ . It can be seen that the diagonal of the square joining the points  $(\phi_1=\pi, \phi_2=0)$  and  $(\phi_1=0, \phi_2=\pi)$  separates the positive from the negative values of  $s$ . In finding the arithmetic mean value we need consider only the positive values in the half of the square in which  $(\phi_1+\phi_2)<\pi$ . If the total number of points in this triangle is  $N'$  there are

$(2N'/\pi^2)d\phi_1 \cdot d\phi_2$  points in a small element of area  $d\phi_1 \cdot d\phi_2$  at  $(\phi_1, \phi_2)$ . For these trials  $s=l(\cos \phi_1 + \cos \phi_2)$  and the arithmetic mean value of  $s$  is

$$\frac{2l}{\pi^2} \int_0^\pi d\phi_1 \int_0^{\pi-\phi_1} (\cos \phi_1 + \cos \phi_2) d\phi_2 = \frac{8l}{\pi^2} = 0.811l. \quad (17)$$

This may be obtained directly from the value of  $P_2(s)$  given in equation (16), since <sup>(8)</sup>

$$\frac{1}{l^2} \int_0^2 sF(\theta) ds = 4 \int_0^1 \cos \theta \cdot F(\theta) \cdot d(\cos \theta) = 4. \quad (18)$$

The probable value of  $s$  cannot be expressed in terms of elementary functions but may be found by measurement of the area under the probability distribution curve. Its value is  $p=0.710l$ . The R.M.S. value of  $s$  is  $l$ .

### 9. *The Solution for $n=3$ .*

It was felt desirable to obtain a solution for  $n=3$ , for two reasons.

- (a) The probable value of  $s$  for  $n=2$  was practically the same as for the normal distribution and it was of interest to determine whether this rapid convergence was maintained in the step to the next higher value of  $n$ .
- (b) It was suspected that the probability function would be finite for all values of  $s$  when  $n=3$  and it was desired to determine whether this was so.

The solution is derived from equations (12) and (16).

$$\begin{aligned} P_3(s) &= \frac{1}{\pi} \int_0^{\cos^{-1}(\frac{s-X}{l})} P_2(s-l \cos \alpha) d\alpha \\ &= \frac{1}{\pi^3 l} \int_0^{\cos^{-1}(\frac{s-X}{l})} F(\theta_1) d\alpha, \quad (19) \end{aligned}$$

where  $F(\theta_1)$  is the complete elliptic integral corresponding to that defined in section 8 and

$$\cos \theta_1 = \frac{s-l \cos \alpha}{2l}. \quad (20)$$

The upper limit of integration depends on the range in which  $s$  lies since  $(s+l)$  may or may not be greater than  $2l$  [see equation (9)]; there is a discontinuity in slope at  $s=l$ .

For  $0 \leq s \leq l$  the upper limit is  $\cos^{-1}(-1)=\pi$

For  $l \leq s \leq 3l$  the upper limit is  $\cos^{-1}(s/l-2)$ .

The integral cannot be expressed in terms of tabulated functions and the following method of evaluation was used.  $F(\theta_1)$  was plotted against  $\alpha$  for each value of  $s$  (mainly multiples of  $0.2l$ ) and the area under the curve measured between the appropriate limits. For  $0 \leq s \leq l$ ,  $F(\theta_1)$  becomes infinite at  $\alpha=\cos^{-1}(s/l)$ , which introduces a difficulty in the measurement of the area under the curve. The difficulty was overcome by finding an expression for the area in the form of an infinite series which could be used for a small range of  $\alpha$  including the value  $\cos^{-1}(s/l)$ . This series was based on the expression quoted by Pearson on page xx. of the introduction to the tables of elliptic integrals <sup>(7)</sup>. The area under the rest of the curve was measured by a standard method involving the summation of ordinates.

In fig. 6,  $lP_3(s)$  is plotted against  $s/l$ . As a check on the accuracy of the calculations the area under this curve was measured and found to be equal to one-half to about 1 part in  $10^4$ , which was well within the estimated accuracy of working.

The case of  $n=3$  is capable of a three-dimensional representation similar to the alternative two-dimensional representation for  $n=2$ . The

Fig. 6.

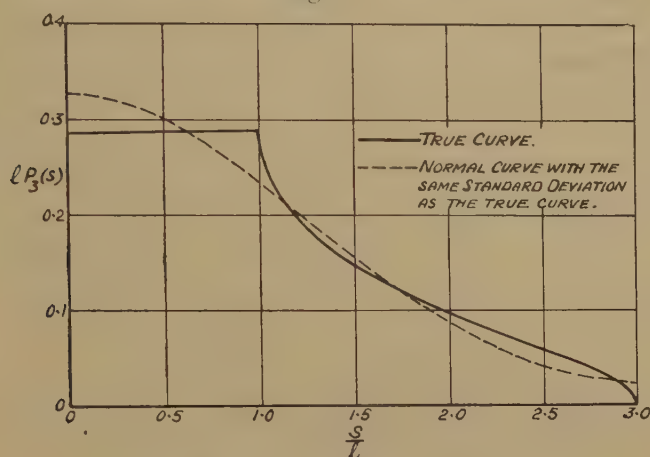
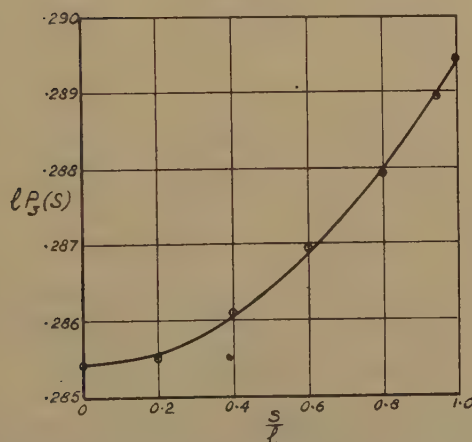
Probability distribution for  $n=3$ .

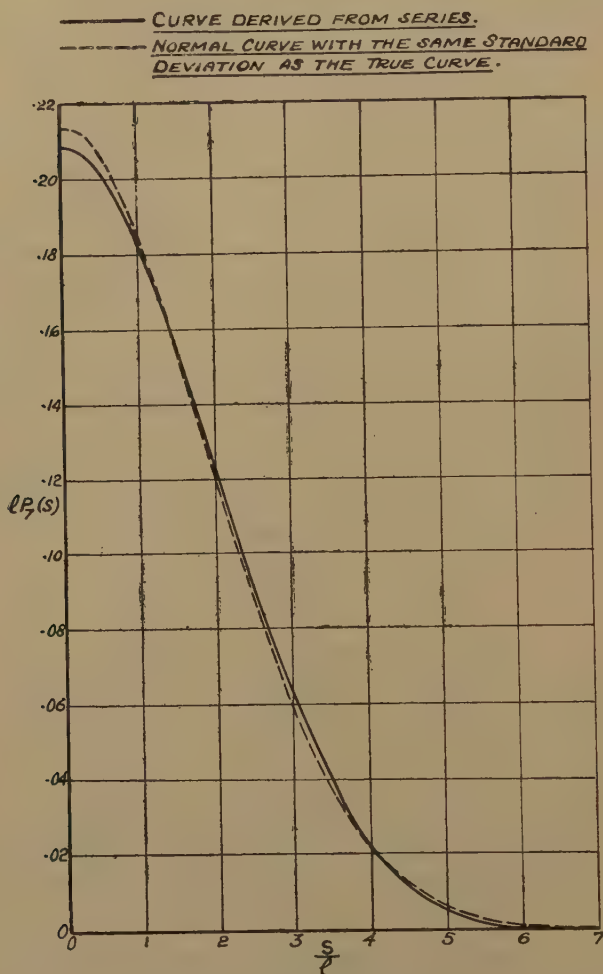
Fig. 7.

Part of probability distribution for  $n=3$   
(section of fig. 6 drawn to enlarged scale).

square of side  $2\pi$ , of which one-quarter is shown in fig. 5, is replaced by a cube of edge  $2\pi$ , the values of  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  being measured along the three mutually perpendicular edges. A large number of points, each representing the result of a selection of  $\phi_1$ ,  $\phi_2$  and  $\phi_3$ , are uniformly distributed throughout the cube and values of  $s=l(\cos \phi_1 + \cos \phi_2 + \cos \phi_3)$  are represented by

curved surfaces within the cube. The use of this alternative approach has not led to a simple expression for the arithmetic mean value of  $s$ , as was the case for  $n=2$ .

Fig. 8.

Probability distribution for  $n=7$ .

The following arithmetic mean and probable values have been determined graphically from the plot of the probability function.

Arithmetic mean value =  $1.001l$ .

Probable value =  $0.873l$ .

The R.M.S. value is  $l\sqrt{3/2} = 1.225l$ .

#### 10. *The Solution for $n=7$ .*

Pearson found that the solution to problem A could be expressed in terms of infinite series and that, for values of  $n$  greater than 6, very few



terms of these series were necessary to yield solutions scarcely distinguishable from those obtained by graphical methods. This suggested that the series might prove useful for solving problem B in similar circumstances. They were, in fact, found adaptable for the purpose and a solution for  $n=7$  was computed by this method using the first six terms of the appropriate series, the same number as used by Pearson.

The resulting curve of  $lP_7(s)$  against  $s/l$  is plotted in fig. 8, and is seen to resemble very closely the normal curve derived from equation (3) which has been plotted for comparison on the same graph. The series naturally cannot represent the true curve accurately in the neighbourhood of  $s=7l$  unless an infinite number of terms are used, since otherwise finite values of  $P_7(s)$  are obtained for values of  $s$  greater than  $7l$ . Nevertheless the series do give values very much smaller than the normal curve for values of  $s$  in the neighbourhood of  $7l$  even when few terms are used. For instance, when  $s=7l$  the normal curve gives  $lP_7(s)=0.0002$ , while six terms of the series give  $lP_7(s)=0.00001$ .

The arithmetic mean and probable values of  $s$  can be derived from Pearson's work and are as follows :—

Arithmetic mean value  $1.513l$ .

Probable value  $1.292l$ .

The R.M.S. value is  $l\sqrt{(7/2)}$  or  $1.871l$ .

#### 11. Discussion of the Results for $n=1, 2, 3$ and $7$ .

The curves of the probability functions are plotted for positive values of  $s$  only, since they are symmetrical about the axis  $s=0$ , and the normal curves derived from equation (3) have been added for comparison. The true curves show the following similar characteristics to Pearson's curves. For purposes of comparison the normal solution of problem B is analogous to the Rayleigh solution of problem A, although this is not strictly of normal form.

1. For small values of  $n$  the curves are widely divergent from the normal form.

2. The probability function becomes infinite for some value of  $s$  when  $n=1$  or  $2$  and the curves for small values of  $n$  greater than  $2$  have discontinuities in slope.

3. The true probability distribution tends progressively towards the normal form as  $n$  is increased.

4. The curve for  $n=7$  is very close to the normal form.

It is evident that the curve of  $P_n(s)$  bears a closer resemblance to the normal form than Pearson's corresponding curve for  $\Phi_n(r^2)$  does to the Rayleigh solution of problem A. There are no infinities in  $P_n(s)$  when  $n=3$ , and it follows, from equation (12), that the functions are always finite for values of  $n$  greater than  $3$ . In contrast the functions  $\Phi_n(r^2)$  have infinities for all values of  $n$  less than  $5$ . There is some resemblance

between  $P_2(s)$  and  $\Phi_4(r^2)$  and a very marked resemblance between  $P_3(s)$  and  $\Phi_5(r^2)$ . Pearson found that  $\Phi_5(r^2)$  was constant for  $0 \leq r \leq l$  within the accuracy of calculation, which appears to have been better than 1 part in 500. A similar characteristic is exhibited by  $P_3(s)$  in the range  $0 \leq s \leq l$ , but this function diverges from constancy to a much greater degree than  $\Phi_5(r^2)$ , as can be seen from fig. 7, which shows a portion of fig. 6 plotted on a larger scale. The scatter of the points about the smooth curve is within the estimated error of computation.

It appears then, that the convergence of the function  $P_n(s)$  towards the normal form, as  $n$  is raised, is more rapid than the convergence of  $\Phi_n(r^2)$  to the Rayleigh solution of problem A. Comparing the present results with those of Pearson it seems safe to assume that a plot of  $P_n(s)$  will be a sensibly smooth curve for  $n=5$ . Since it has the same area and standard deviation as the normal curve, it also seems reasonable to suppose that the true curve will be closely represented by the normal expression (3) for  $n=5$ . The computation of the probability functions for  $n=4, 5$  and  $6$ ,

TABLE I.

$n$	Probable Value		Arithmetic Mean Value		R.M.S. Value
	True	Normal	True	Normal	
1	0.707	0.477	0.637	0.565	0.707
2	0.710	0.674	0.811	0.798	1.000
3	0.873	0.826	1.001	0.978	1.225
7	1.292	1.262	1.513	1.493	1.871

which would be very laborious, has not, therefore, been considered justifiable.

The results for  $n=1, 2, 3$  and  $7$  are summarised in Table I., in which the arithmetic mean and probable values of  $n$  given by the normal expression (3) are compared with the values derived from the more exact analyses. The R.M.S. value of  $s$ , which is the same for both distributions, is also given, and  $l$  has been taken as unity throughout the table, for convenience. It should be borne in mind that the values for the normal curve are derived from the whole of the curve and not from that part only which lies between the limits  $s=\pm nl$ .

Examination of this table shows that the rapid concurrency of the tabulated features of the true and normal distributions between  $n=1$  and  $n=2$  is not maintained in the step from  $n=2$  to  $n=3$ . Nevertheless, the difference between the two distributions in respect of the tabulated features is small at  $n=2, 3$  and  $7$ , and there is no reason to suppose that there will be any large increase in this difference at any value of  $n$  between  $3$  and  $7$  or higher than  $7$ .

## 12. Conclusions.

The paper has shown that the complete solution to the problem under investigation can be formulated mathematically without difficulty, but, owing to the complexity of the functions involved, the derivation of the probability distributions is very laborious. Fortunately the true probability distributions converge rapidly towards the normal form as  $n$  is raised. The solution to the problem is expressed in terms of the probability function  $P_n(s)$ , where  $P_n(s)ds$  expresses the probability that the amplitude of that component of the resultant simple harmonic function with the chosen phase will lie between the limits  $s$  and  $s+ds$ . Curves have been given showing the variation of  $lP_n(s)$  with  $s/l$  for  $n=1, 2, 3$  and  $7$ , where  $l$  is the amplitude of the individual harmonic functions. The curve for  $n=3$  was derived by graphical means and that for  $n=7$  by an infinite series which can be used, with very little labour, to derive curves for any value of  $n$  greater than  $7$ .

There are no infinities in the probability function for values of  $n$  greater than  $2$ , and it seems probable that a plot of the function will be a sensibly smooth curve for any value of  $n$  greater than  $4$ . As  $n$  is increased the probability function tends progressively towards the normal form having the correct standard deviation and the normal expression certainly gives a close representation of the true facts when  $n=7$ , for values of  $s$  between the limits  $\pm nl$  and probably does so for  $n=5$  and  $n=6$ . The probability function is necessarily zero for values of  $s$  outside the limits  $\pm nl$ .

The R.M.S. value of  $s$  is the same for the true and the normal distributions when the whole of the normal curve is taken into consideration. Its value is  $l\sqrt{(n/2)}$ . The arithmetic mean and probable values are almost the same for both distributions for all values of  $n$  except unity, and for practical purposes it may be assumed that these values are  $0.564\ l\sqrt{n}$  and  $0.477\ l\sqrt{n}$  respectively within a few per cent.

These results are applicable to the study of the site error when  $n$  interfering rays of equal amplitude arrive at a direction-finder from a narrow range of azimuth, and the phases of these are completely random. It is deduced that there is a 50 per cent probability that the resultant error will be less in magnitude than  $0.477\ l\sqrt{n}$  times the maximum error which could be caused by any individual ray, provided that the maximum possible error due to all the rays is not large. This is a close approximation even when  $n$  is only  $2$ . Similarly, if the individual phases are continually changing at random over a range  $0$  to  $2\pi$ , the error will be less than the above value for half the time. If a large number of errors due to  $n$  interfering rays of equal amplitude but continually varying phases are measured, the R.M.S. error will be  $\epsilon_{\max}\sqrt{(n/2)}$  and the arithmetic mean error  $0.564\ \epsilon_{\max}\sqrt{n}$ , where  $\epsilon_{\max}$  is the maximum error which could be caused by any one of the interfering rays alone.

The arrival of several interfering rays of equal amplitude from a small range of azimuth is a special case of the more general condition to which the theory is applicable, that the amplitude of the pick-up on the direction



finder is the same from all the rays. This condition may be satisfied when the rays are unequal and arrive from several different azimuths, but in applying the theory to determine the probable resultant error it will, in general, be necessary to assume that the error is small and that the pick-up from each individual ray is sensibly independent of small displacements of the aerial system from its "on-bearing" position.

#### References.

- (1) K. Pearson, "The problem of the Random Walk," 'Nature,' lxxii. p. 294 (1905).
- (2) K. Pearson, "Mathematical Contributions to the Theory of Evolution XV. A Mathematical Theory of Random Migration," Draper's Co. Research Memoirs, Biometric Series III, Dulau & Co. (1906).
- (3) Lord Rayleigh, "On the Resultant of a large number of Vibrations of the same Pitch and Arbitrary Phase," Scientific Papers, I, p. 491.
- (4) Lord Rayleigh, "On the Problem of Random Vibrations and of Random Flights in 1, 2 or 3 Dimensions," Scientific Papers, VI, p. 604.
- (5) J. C. Kluyver, "A Local Probability Problem," Proc. Amsterdam Academy of Sciences, viii. pp. 341-350 (1905).
- (6) A. Cayley, 'An Elementary Treatise on Elliptic Functions,' Bell, London, Chapter XII. (1895).
- (7) Legendre-Pearson, 'Tables of the Complete and Incomplete Elliptic Integrals,' Cambridge University Press (1934).
- (8) Whittaker and Watson, 'A Course of Modern Analysis,' Cambridge University Press, p. 521 (1940).

### XX. A Magnetic Lens of Special Field Form for $\beta$ - and $\gamma$ -Ray Investigations, Designs and Applications.

By KAI SIEGBAHN\*.

[Received April 29, 1946.]

#### THEORY.

The possibility of obtaining an electron lens without spherical aberration has previously been discussed in the theoretical literature of electron optics. It has been found that the error due to spherical aberration is always positive. Theoretically, however, the error may be made infinitely small by a suitable choice of the field form and a correct placing of the object. It has not yet been elucidated how far the spherical error may be eliminated in practice in view of the field form actually realizable and other factors. According to Rebsch † the least aberration error  $\rho$  (defined as the radius of the confusion disk) may be estimated as

$$\rho \geq \frac{1}{4} f V \delta^3, \quad . . . . . (1)$$

where  $V$  is the magnification,  $\delta$  the aperture of the objective and  $f$  the focal distance. The lenses hitherto used in electron-microscopy have appreciably greater aperture errors than the limit set by (1) ‡.

\* Communicated by Sir George Thomson, F.R.S.

† R. Rebsch, *Ann. d. Phys.* v. (1938).

‡ M. v. Ardenne, 'Elektronen-Übermikroskopie,' Berlin, 1940, p. 45.

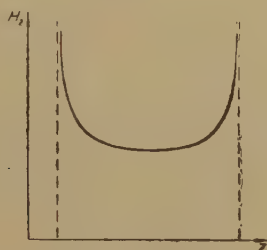


Numerical calculations of the lens errors by Glaser \*, however, show that the limit for the spherical error should be placed considerably lower than that imposed by (1). In one of his papers, Glaser † has even calculated a field form that should give a zero spherical-aberration error. He starts from the formula for the spherical error derived in electron optics ‡.

$$\rho = V \cdot \alpha^3 \cdot \frac{e}{96mU} \int_{z_1}^{z_2} \left( \frac{2e}{mU} H^4 + 5H'^2 - HH'' \right) h^4 \cdot dz, \quad . \quad . \quad . \quad (2)$$

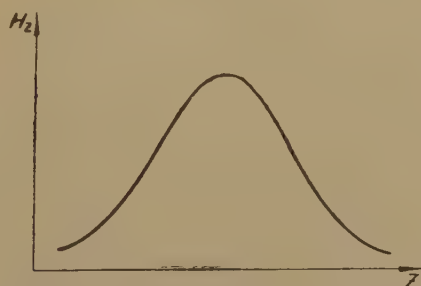
where  $\alpha$  denotes the aperture,  $e/m$  the specific charge of the electron,  $U$  the acceleration potential,  $h=r/r_B$ ,  $r_B$  the distance of the electron

Fig. 1.



Field form according to Glaser, giving a spherical error = 0.

Fig. 2.



Bell-shaped field, commonly used in magnetic lenses.

from the axis at  $z=z_B$ . By putting the integrand equal to zero a differential equation is obtained, the solution of which gives the field form corresponding to zero aberration error. It is found that the field varies in the manner shown in fig. 1. As at two points the field strength increases to infinity, this field form can obviously not be realized in practice. Glaser finds, however, that an almost negligible error is introduced if the field strength is given finite values at these points. The ideal field form given by Glaser as free from spherical aberration, can nevertheless lead

\* W. Glaser, *Z. f. Phys.* cxvi. p. 734 (1940); cxvii. p. 285 (1941).

† W. Glaser, *Z. f. Phys.* cxvi. p. 19 (1940).

‡ W. Glaser, *Z. f. Phys.* cix. p. 700 (1938).

to other difficulties, namely, that the refractive power of the lens is so strongly diminished that no image is formed\*.

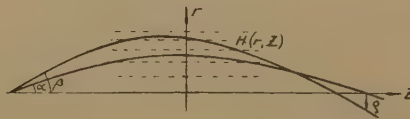
This field form clearly differs in essential respects from the bell-shaped type usual in electron-microscopy (fig. 2) and gives ample reason for a closer experimental study. Unfortunately it is impossible to achieve the Glaser field form in its entirety. However, there are no practical difficulties in obtaining a field with the characteristic convexity.

A simple consideration already shows that this field form should be suited to reducing the spherical error. Let us assume that the field strength along the axis of the lens is  $H(z)$ . It results† that the axial component of the field strength throughout the space  $H_z(r, z)$  can be derived from  $H(z)$  and its derivatives according to

$$H_z(r, z) = H(z) - \frac{r^2}{4} H''(z) + \dots \quad (3)$$

We shall now investigate what factors influence the spherical error in the image of a point on the axis of a magnetic lens. In fig. 3 are shown two different electron tracks passing from the object at different angles  $\alpha$  and  $\beta$  with relation to the lens axis  $z$ , which is at the same time the axis of symmetry of the magnetic field  $H(r, z)$ . The electron tracks,

Fig. 3.



Schematic drawing showing the spherical aberration.

which are subjected in the lens to a deflection around the  $z$ -axis, are developed in the  $r$ - $z$ -plane in the figure.

The electron track that forms the larger angle  $\beta$  with the axis, will cut the latter earlier than the track with the smaller angle ( $\alpha$ ), since its velocity projection along the axis is smaller. We obtain, in consequence, a confusion disk in the plane of the image, the radius of which is denoted in the figure by  $\rho$ . This spherical aberration is always positive by definition.

The spherical aberration error, however, includes another spherical image error independent of that mentioned above. This is due to the gradient of the  $z$ -component of the field strength in the  $r$ -direction. If this gradient is positive, an electron passing from the axis at a large angle will move in a field which is stronger than that acting upon an electron track nearer the axis. In this case we have a positive addition to the spherical error. It may be stated as a general rule that the bell-shaped field of the type shown in fig. 2 has this characteristic.

With the aid of equation (2), however, it is easy to deduce the appear-

\* R. Rebsch, *Z. f. Phys.* cxvi. p. 729 (1940).

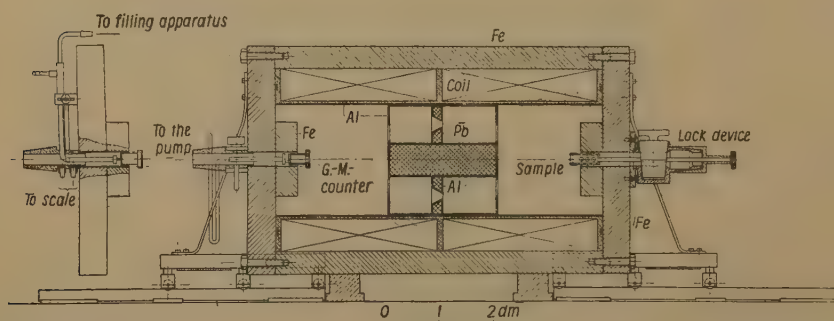
† Cf. K. Siegbahn, *Ark. Mat. Astr. Fysik.* xxxA. No. 1 (1943).



preliminary experiments, however, their dimensions were fixed at 13 cm. diameter and 4 cm. thickness. The diameter of the holes is 2.8 cm.

The GM tube is placed before the hole in one of the end-pieces. It is connected by a tube with the filling apparatus for the appropriate gas mixture. Behind the GM tube the evacuation tube is also fitted to the end-piece of the lens by means of a conical joint. When the spectrograph is being evacuated the GM tube is in connection with its interior through a tap, so that the same pressure will prevail on both sides of the thin foil on the GM tube. By closing the tap, the argon-alcohol mixture (90+10 mm.), contained in a large glass bottle, may be placed in connection with the GM tube. Thanks to this device the foil on the GM tube may be made much thinner than if it were to stand atmospheric pressure. The amplifier equipment of the GM tube comprises a Neher-Harper extinguishing circuit, a scale of eight of a somewhat modified Lifschutz type,\* and a mechanical counter of high resolving power

Fig. 5.



Plan of the spectrograph.

( $\tau=1/600$  sec.). The high-tension unit for the GM tube consists of a Street-Johnson circuit†.

The sample is introduced into the vacuum by means of a lock device. This is of importance not only in the study of short-lived samples. In point of fact, it has usually proved necessary to allow the GM tube to stabilize its working plateau for about half an hour before the experiment can commence. If the sample-lock were absent, the spectrograph would have to be filled with air when the sample was introduced, which the windows of the GM tube would not tolerate.

The shutter system is placed between the GM tube and the sample. It is mounted in an aluminium tube which can readily be moved out from the spectrograph. An axially mounted cylinder of solid lead, of 2 dm. length and covered with aluminium, effectively prevents direct  $\gamma$ -radiation from the sample from reaching the GM tube. The pencil of rays is limited by three circular shutters of lead, which are also coated

\* Lifschutz, *Rev. Scient. Instr.* x. p. 21 (1939).

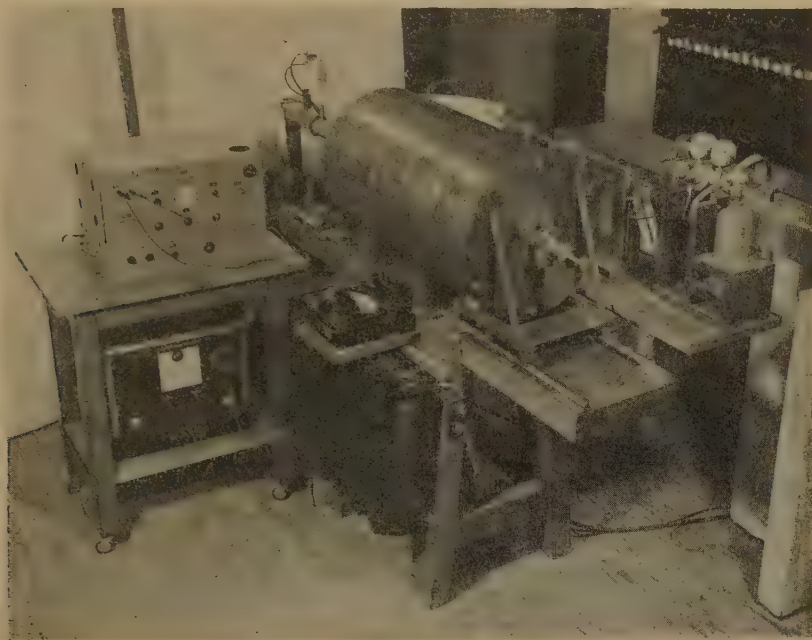
† Street, J. and Johnson, T., *Franklin Inst.* ccciv. p. 155 (1932).



with aluminium to reduce the secondary radiation. All shutters are easily exchangeable, so that the resolving power of the spectrograph may be increased or reduced if required.

Fig. 6 is a photograph of the whole installation. The sample-lock may be seen on the forward end-piece of the spectrograph. Behind the spectrograph is the filling apparatus for the GM tube. Centrally situated before the spectrograph is the resistance for regulating the current through the lens, mounted together with an accurate ammeter. On a special table to the left is the amplifier for the GM tube.

Fig. 6.



Total view of the apparatus.

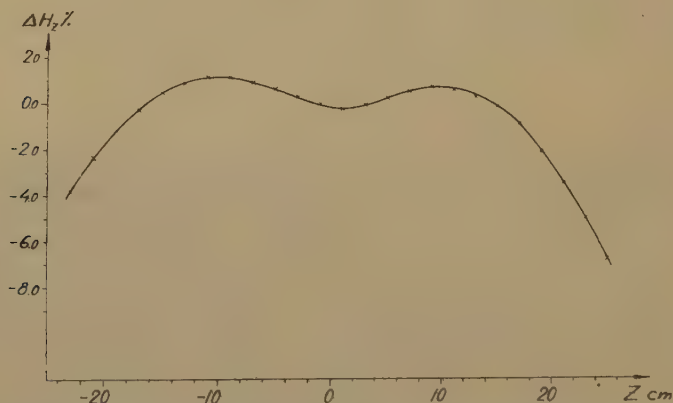
#### MEASUREMENT OF THE MAGNETIC FIELD.

According to equation (3) it is sufficient to measure the field strength along the axis in order to know its value throughout the space. It should, however, also be of interest to measure  $\delta H_z / \delta r$  experimentally at various points in the lens, since it is this quantity that determines the contribution from the field to the spherical aberration error.

The measurement was performed in the usual manner with a fluxmeter and a search-coil which could be moved about in the field with the aid of a mechanical device. For the purpose of finding the influence of the pole-pieces on the field form, the field was determined in both the presence and the absence of these. Several different sizes and forms of pole-piece were investigated. Here, however, only those results have been included which concern the ultimately selected size ( $d=4$  cm.,  $\phi=13$  cm.).

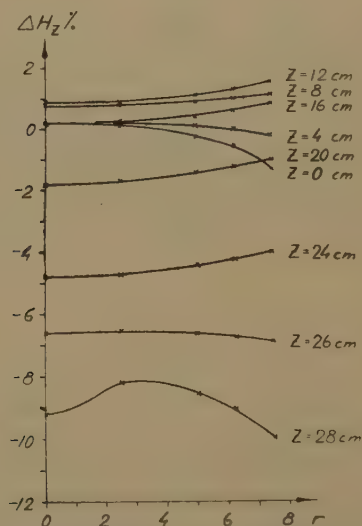
Fig. 7 shows the variation in the field strength along the axis in the absence of pole-pieces. Within a very wide range in the centre of the lens the field strength is apparently quite constant with a weak minimum in the middle. In the neighbourhood of the end-pieces the field shows a marked weakening.

Fig. 7.



The magnetic field along the lens axis in the absence of pole-pieces.

Fig. 8.



The variation of the magnetic field along the radius in the absence of pole-pieces.

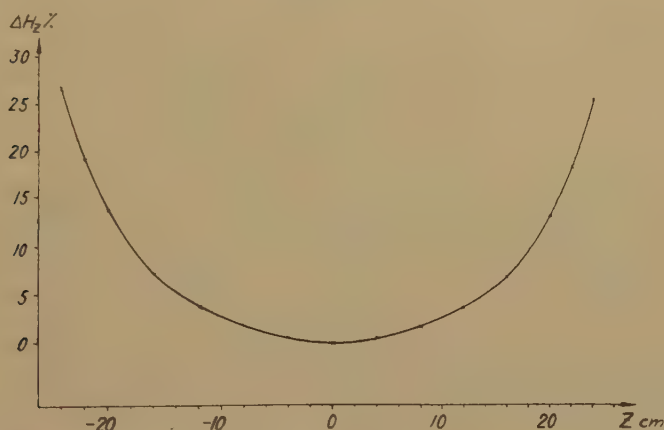
In view of the fact that, according to fig. 7, the second derivative in the centre of the lens is positive, a negative gradient  $\delta H_z / \delta r$  is to be expected in this region, afterwards reversing to the positive sign. This agrees quite well with the experimental recording of the variation in  $H_z$  at different points on the radius in the lens of fig. 8.  $z$  denotes here

the distance from the centre of the lens. Not before  $z=28$  is a small effect noticed, due to the boring in the end-piece.

The introduction of pole-pieces has a marked effect on the field in the desired direction. This is shown in fig. 9. The downward convex field form required has here been achieved. It should, however, be pointed out at this point that, by another choice of the field-determining factors (the length and diameter of the lens and the pole-pieces) it should doubtless be possible to obtain fields that have more favourable influence on the spherical aberration than that attained here.

Since the field along the axis in the whole region has a downward convex form, the derivative  $\delta H_z/\delta r$  should, according to (3), have the desired negative sign at all points. Fig. 10 shows that this is actually the case.

Fig. 9.



Magnetic field along the lens axis with pole-pieces.

#### RESOLVING POWER AND LIGHT INTENSITY.

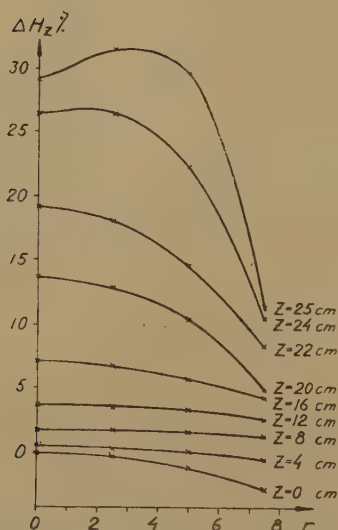
For the study of the spherical error it is fitting to investigate the resolving power and light intensity of the spectrograph for different shutter apertures. The F-line of ThB has therefore been studied in four different cases: with aperture A,  $\phi$  external=14,  $\phi$  internal=12 cm.; aperture B,  $\phi_e=14$ ,  $\phi_i=10$  cm., in which cases the diameters of the sample and the GM tube slit were 4 mm. and with aperture C,  $\phi_e=14$ ,  $\phi_i=12$  cm.; and aperture D,  $\phi_e=14$ ,  $\phi_i=10$  cm., the diameters of the sample and the GM tube slit being 8 mm. in the latter two cases. The result is shown in fig. 11, the strength of the sample in each case being reduced to the same magnitude.

The presence of the spherical error is manifested by the difference in the values of the current required to focus the F-line in cases A and B on the one hand and C and D on the other.

For samples with high specific activity the spectrograph should be adjusted for cases A and B. If high resolving power is especially

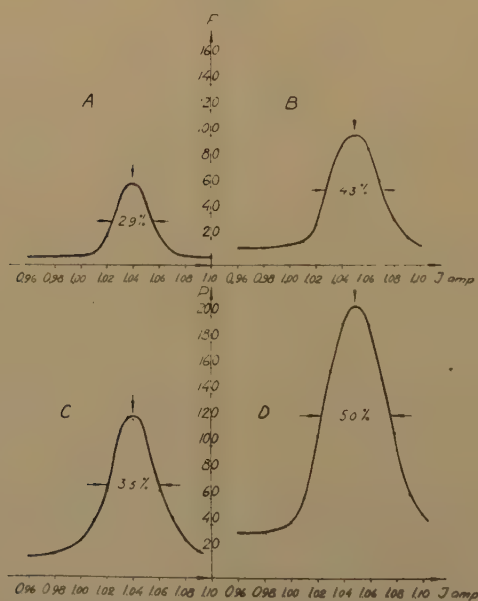
desirable, A is to be preferred, whereas B gives about double the light-intensity.

Fig. 10.



The variation of the magnetic field along the radius with pole-pieces.

Fig. 11.



The resolving power in different cases.

In the great majority of cases of artificial radio-activity, samples of relatively low specific activity are concerned. Cases C and D should



therefore have the most extensive application. Especially in the study of continuous spectra and investigations of  $\gamma$ -rays by secondary radiation, case D is without doubt the most suitable. (See below : Applications.)

An approximate estimation of the space-angle used in the case of highest light-intensity with  $\phi_e=14$  and  $\phi_i=10$  cm. is obtained by the following considerations :—

In a homogeneous magnetic field electrons move in helical paths. When projected on a single plane this movement becomes a sine curve (fig. 12). The initial directions of the electrons will be bounded in this case by the tangents to the two sine curves passing through the middle shutter of the spectrograph with 5 and 7 cm. radii. The distance between

Fig. 12.



Diagram for computing the approximate light-intensity.

the sample and the slit of the GM tube is 50 cm. The equations for the two outermost rays will be

$$r=5 \cdot \sin \left( z \cdot \frac{\pi}{2} \cdot \frac{1}{25} \right),$$

$$r=7 \cdot \sin \left( z \cdot \frac{\pi}{2} \cdot \frac{1}{25} \right),$$

whence we obtain

$$\tan \phi_1 = \frac{5\pi}{50},$$

$$\tan \phi_2 = \frac{7\pi}{50},$$

i. e.,

$$\phi_1=17^\circ 26' \text{ and } \phi_2=23^\circ 44'.$$

If the magnetic field were homogeneous this space-angle for the electron beam would correspond to about 2 per cent. of the total radiation passing the shutter system. As, however, the field in the lens described here is the stronger, the shorter the distance from the sample and the GM tube, we can count with a still greater initial angle for the electron beam with a consequent increase of the space-angle utilized. It is possible to calculate theoretically the space-angle corresponding to the field under consideration here, although the calculation is complicated by the spherical aberration error occurring simultaneously.

It may be of interest to compare the light-intensity obtained with this lens at a given resolving power with that from a lens with a bell-shaped field. The present author has earlier constructed and used an example of the latter \*. The length and internal diameters of the coil of this lens are both 30 cm. The distance between the GM tube and the sample is 60 cm. The lens has a bell-shaped field, which may be described analytically by  $H(z) = H_0 \cdot e^{-\left(\frac{z}{b}\right)^2}$ , and is upwardly convex within a range of about 36 cm. in the centre of the lens. A direct experimental investigation now shows that the new lens possesses, for a given resolving power, double the light-intensity of the old instrument. As has earlier been indicated, a still more favourable choice of a downward convex field form would probably further improve the image-forming properties of a lens of this type.

#### DISCUSSION OF THE APPLICABILITY OF THE SPECTROGRAPH.

The advantage of the lens method over the semi-circular method naturally lies above all in the fact that the focusing is obtained in all planes and not only in that perpendicular to the lines of force. Moreover, by choosing a suitable field form, which reduces the spherical error as has been done in the present case, a considerable space-angle can be employed. Another considerable advantage is the large distance between the sample and the GM tube, which makes possible the effective screening-off unwanted radiation. This is especially valuable in the study of  $\gamma$ -spectra.

The "open" magnetic lens constructed by the author has certain advantages over the "closed" lens, if the lower light-intensity be disregarded. This especially applies to the accessibility of the sample and the GM tube, which may both be placed completely outside the lens. On the other hand, the closed model provides a very effective screening from all external fields, such as the magnetic field of the earth. It is, hence, unnecessary to provide special compensation coils to correct for the disturbances caused by the latter.

All spectrographs constructed with iron have the disadvantage that the linear relation between current and field strength is not quite valid. By selecting iron with low remanence, however, this unfavourable condition can largely be removed. Nevertheless, it is advisable to demagnetize the lens with alternating current after each occasion when it has been employed for focusing spectra of high energy. With the above-described lens the author has been unable to establish any departure from the linearity between the current and the focused  $H\rho$ , apart from a small effect at very high energies ( $=4$  MeV.), where the correction for this is due to an incipient tendency to saturation amounts to only one or two per cent. At extremely low energies a small deviation is also to be expected, especially if care has not been taken to demagnetize

---

\* Kai Siegbahn, *Ark. f. Mat., Astr. o. Fys.* xxx. A. No. 1 (1943) and No. 20 (1944).

the instrument, since any remanence will have a relatively great influence here. In such cases it is always possible to make a direct measurement of the field strength.

It may be of interest in this connection to point out that magnetic fields where the electrons move at right angles to the lines of force, can also be made to focus the electrons in both the horizontal and vertical planes. Such a possibility is associated with a field where the field-strength diminishes from the centre as  $1/\sqrt{r}$ . Here the focusing does not occur, as in the homogeneous magnetic field, after  $180^\circ$ , but after  $\pi\sqrt{2}$  radians, i. e.,  $256^\circ$ . The focusing properties of a field of this form are at present the object of an investigation by Dr. Svartholm and the present author. Our results will be published at a later date.

#### APPLICATIONS.

##### *General Considerations.*

The study of the artificial radio-active isotopes is at present one of the most important topics in nuclear research. It involves, in the first place, the establishment of a detailed term scheme for the disintegration in question. Here, not only the energies of the different  $\beta$ - and  $\gamma$ -components appearing are of interest, but also their intensities. These latter data give valuable information concerning the probabilities of transition between the different nuclear levels. From these it is already possible to draw conclusions regarding the mechanism of the disintegration, to predict the spin and parity of the levels, etc. Unfortunately the theoretical basis for such considerations is at present only in a preliminary stage. The development in the theoretical field is hindered, among other things, by the fact that the term schemes for only a few disintegrations are yet established experimentally.

The reason for this state of affairs is to be found in difficulties experienced with the intensities in the relevant determinations. Only in exceptional cases are samples obtained with the strengths available in the naturally active isotopes. The specific activities, moreover, are considerably lower in the artificial isotopes, which gives rise to further difficulties.

The study of the  $\gamma$ -radiation of the naturally active isotopes presents no special difficulties, since the atomic numbers of these are sufficiently high to give an internal photo-effect of the  $\gamma$ -lines, so that these are observed in the form of well-defined  $\beta$ -lines. With low atomic numbers of artificially active isotopes this phenomenon occurs only in exceptional cases, namely, when multipole radiation is concerned. It is therefore essential to develop a technique that makes possible, even in the absence of an internal photo-effect, the determination of the energies and intensities of the  $\gamma$ -radiation.

In consequence of the difficulties outlined above, especially in what concerns the intensity, it has been necessary to resort, in a very large number of cases, to absorption methods and investigation with the



Wilson chamber. In some instances the latter technique appears to be especially indispensable, namely, where the disintegrations have extremely short half-life periods. It must, however, be considered highly desirable to replace these methods by the more reliable spectroscopic technique to the greatest possible extent. A presupposition for this is clearly that the light-intensity in the spectroscopic procedure be made sufficiently high to enable us to study samples with activities of about the same order of magnitude as those accessible to the other methods. An account will be given below of the method which the present author has developed in view of these considerations.

With the aid of the lens of high light-intensity described above, it has proved possible to study continuous  $\beta$ -spectra which, in favourable cases, need amount to only some hundredths of a microcurie. Moreover, with the same lens and a special technique, it is possible to investigate  $\gamma$ -spectra whose activity corresponds to only about  $1\mu\text{C}$  Ray-equivalence per  $\gamma$ -line. These strengths do not differ essentially from those required for moderately good absorption measurements. Simultaneously, the spectroscopic procedure permits measurements of an accuracy impossible to achieve, for various reasons, with the two other methods mentioned.

### *$\beta$ -Spectra.*

The study of the  $\beta$ -spectra is highly complicated by the secondary radiation effects which always appear, to a smaller or greater extent, in the backing of the sample, the sample itself, the shutters, etc. These effects are especially disturbing in samples of low specific activity, in which cases the secondary radiation in the sample is responsible for the greater part of the effect. This is especially manifested in the fact that the number of low-energy electrons increases relatively to the remainder of the spectrum. The phenomenon is frequently so marked that simple spectra have occasionally been falsely interpreted as made up of several components.

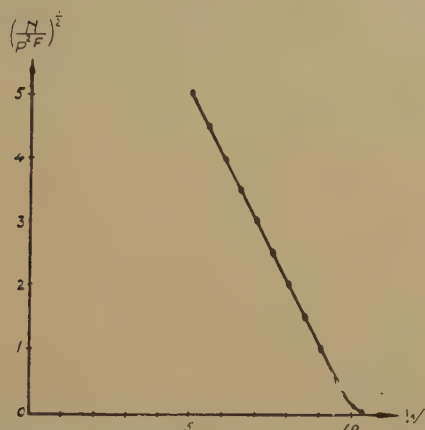
The best precautionary measure against these disturbing effects is to use the smallest possible quantities of substance (requiring high light-intensity in the spectrograph) and to spread the sample over as large an area as possible. Furthermore, thin backing and covering foils must be used for the sample, and the secondary radiation from the inner walls of the spectrograph, and above all the shutters, must be reduced by covering them with an element of low atomic number (*e.g.*, carbon or aluminium).

The two first-named measures require that the spectrograph be adjusted for both favourable light-intensity (small quantity of substance) and moderate resolution (high extension of the sample). In the recording of continuous spectra it is unnecessary to work with high resolving power. Fig. 13 shows the Fermi curve theoretically obtained for a simple spectrum if the spectrograph is set for a resolving power of 5 per cent. (case D in fig. 11). The Fermi curve obtained for infinite resolving power (straight line) actually coincides with this except in its very last



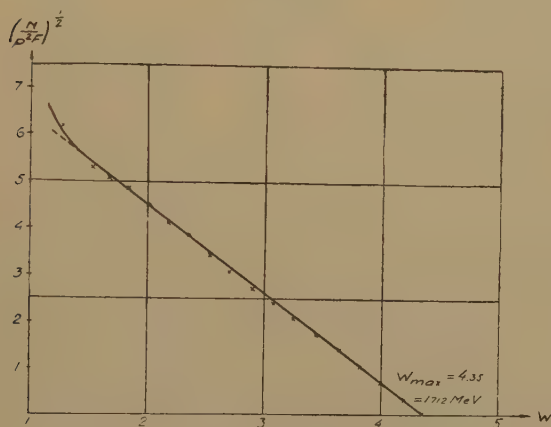
part. If no importance be attributed to the last points in the Fermi plot it is apparently unnecessary to introduce a correction for the finite resolving power. This, however, is obviously not the case if it is desired, for example, to determine the energy maximum by a direct study of the end of the original spectrum. Since this latter procedure lacks accuracy for other reasons (low intensity), the  $\beta$ -spectrum should always be analysed with the aid of the total Fermi plot for the spectrum.

Fig. 13.



Theoretical Fermi curve for a resolving power of 5 per cent.

Fig. 14.

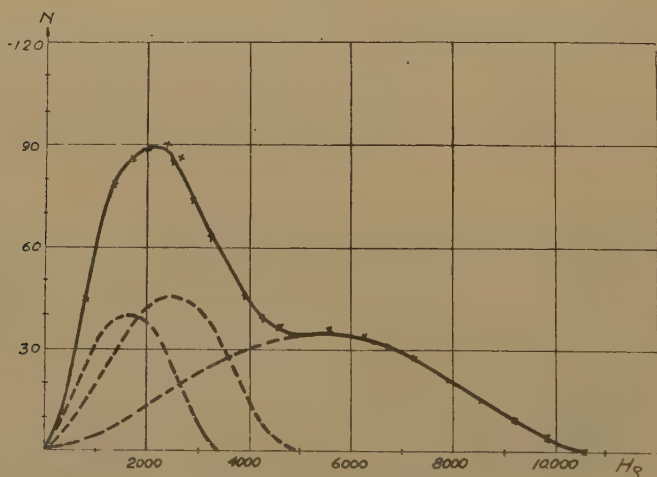

 Fermi diagram for  $P^{32}$ .

It follows from the  $\beta$ -theory that only so-called allowed  $\beta$ -transitions (allowed change of spin=0 or 1) give  $\beta$ -spectra showing straight lines in a Fermi plot. It becomes increasingly clear, however, from the author's recordings of continuous  $\beta$ -spectra, that the greater is the success experienced in eliminating secondary radiation (especially from the sample itself), the more accurately do even the forbidden  $\beta$ -spectra

coincide with straight lines in the Fermi analysis. This will be seen from fig. 14, which represents the Fermi plot for the spectrum of  $P^{32}$ .<sup>\*</sup> According to our conception of this spectrum, it should be forbidden of the second order. Nevertheless, the curve shows excellent correspondence with the straight line. The insignificant divergence at very low energies is doubtless due to a weak residual effect of secondary radiation. The "allowed" form of the spectrum cannot, unfortunately, be made to concord with the  $\beta$ -theory worked out for forbidden transitions<sup>†</sup>.

The experimental fact that both allowed and forbidden  $\beta$ -spectra may be represented as straight lines in the Fermi analysis, highly facilitates the resolution of complex spectra into their components. As an example of this may be cited the disintegration of  $Mn^{56}$  into  $Fe^{56}$  with  $\beta$ - and  $\gamma$ -emission<sup>‡</sup>. Fig. 15 shows the  $\beta$ -spectrum concerned.

Fig. 15.

The  $\beta$ -spectrum of  $Mn^{56}$ .

The Fermi analysis (fig. 16) shows that the original spectrum may be resolved into three components (the broken curves in fig. 15).

It should, however, be pointed out that a resolution of a  $\beta$ -spectrum into its components may often be quite awkward to perform. This is especially the case when several components are to be counted with and when the components of lower energy are low in intensity. In such cases the original Fermi curve gives only uncertain indications of how the spectrum is to be resolved.

The only reliable way to analyse the spectrum is then to investigate the  $\gamma$ -radiation of the disintegration. Since the differences in energy between the  $\beta$ -components must be obtained in the form of  $\gamma$ -lines,

<sup>\*</sup> Kai Siegbahn, Phys. Rev. (1946), in press.

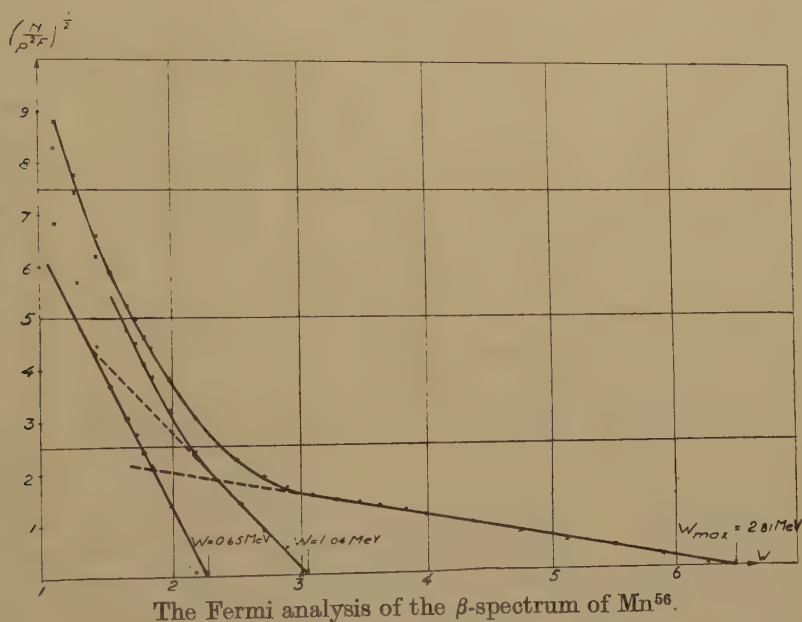
<sup>†</sup> E. J. Konopinsky and G. E. Uhlenbeck, Phys. Rev. lx. p. 308 (1941); cf. also E. J. Konopinski, "Beta-decay," Rev. Mod. Physics, xv. p. 209 (1943).

<sup>‡</sup> Kai Siegbahn, Ark. f. Mat., Astro. o. Fys. xxxiii. A. No. 10 (1945)

the energies of the latter, when known, may be employed in the Fermi analysis.

It is also necessary to study the  $\gamma$ -radiation for other reasons. It frequently occurs that  $\beta$ -transitions to different excite-levels in the final nucleus are so strongly forbidden by the selection rules that they are completely lacking. Even the  $\beta$ -transition to the ground-level in the final nucleus may be completely forbidden ( $\text{Na}^{24}$ ,  $\text{Mn}^{56}$ ), so that this final state can only be attained by  $\gamma$ -emission. Especially in such case is it possible to obtain a completely erroneous conception of the term scheme for the disintegration, if the study of the  $\gamma$ -spectrum is neglected.

Fig. 16.



### $\gamma$ -spectra.

Matters are complicated here by difficulties of intensity to an even higher degree than for the  $\beta$ -spectra. The method hitherto most frequently employed has therefore been the absorption technique. The accuracy thus obtainable is, however, low, and the method is limited to the case where there is only one  $\gamma$ -line, or perhaps a small number with high differences of energy.

For samples of relatively good intensity other methods have found employment. Of these special mention should be made of the use of the nuclear photo-effect, and pair-formation which can be studied in the Wilson chamber. Both these methods, however, are associated with the difficulty that only  $\gamma$ -lines of sufficiently high energies are accessible. For pair-formation the limit is, as is well known,  $E_\gamma > 2mc^2$ , and for the nuclear photo-effect  $E_\gamma = 1.62$  MeV if the reaction  $\text{Be}^9(\gamma, n) \text{Be}^8$  is

employed and  $E_\gamma = 2.18$  MeV for the reaction  $D(\gamma, n)H_1^1$ . The more direct method of determining  $\gamma$ -energies by crystal-reflection is only applicable at relatively low energies and when the strength of the available sample is extremely high.

A much wider applicability has been shown by the methods based on the study of the secondary electrons generated by the  $\gamma$ -radiation as a consequence of the photo and Compton effects in different materials. The most common procedure is presumably to place a foil of aluminium (Compton effect) or lead (photo and Compton effects) in a Wilson chamber, whereafter the  $\gamma$ -radiation from the sample (placed at some distance from the foil), after passing through suitable shutters within a certain limited angular zone, is allowed to strike the foil. The  $\gamma$ -radiation from artificial radio-active isotopes has been studied in this manner by, *inter alia*, Richardson and Kurie\*.

In order to make even weak samples accessible for  $\gamma$ -investigations, Mandeville† and Curran, Dee and Strothers‡ placed the sample in direct contact with an extended aluminium plate, serving as a source of secondary radiation, the thickness of which was sufficient to absorb completely the primary continuous  $\beta$ -spectrum. The secondary electrons were registered in a  $\beta$ -spectrograph of the semi-circular type. The spectrograph was in direct connection with three GM tubes in coincidence coupling, so that the same pressure prevailed throughout. Mandeville reports 0.5 MeV to be the lowest limit for the energy of the  $\gamma$ -radiation that can be studied by this method.

Investigations by Deutsch, Elliott and Evans§ and by the present author show that the lens method has very great advantages in the study of  $\gamma$ -radiation. The large distance of the GM tube from the sample makes possible a quite effective protection of the tube from stray radiation and  $\gamma$ -radiation from the sample. Thus, no coincidence coupling is necessary, but the lens may be used in the same manner as in the registration of  $\beta$ -spectra. The greatly reduced spherical error in the author's lens makes possible the study of samples with strengths as low as  $1\mu\text{C Ra}$   $\gamma$ -equivalence per  $\gamma$ -line.

The procedure may be outlined as follows. The active sample is pressed into a small copper cylinder, the walls of which are thick enough to absorb the continuous  $\beta$ -spectrum. Owing to the high density of copper the radiator may be made small. For a given diameter of the radiator, say 8 mm., a larger active quantity of substance can thus be pressed in than if the radiator were made of aluminium. At energies higher than 0.5 MeV almost only Compton effect is obtained. At

\* J. R. Richardson and F. N. D. Kurie, Phys. Rev. l. p. 999 (1936).

† C. E. Mandeville, Phys. Rev. lxii. p. 309 (1942).

‡ S. C. Curran, P. I. Dee and J. E. Strothers, Proc. Roy. Soc. 1744, p. 546 (1940).

§ M. Deutsch, L. G. Elliott and R. D. Evans, Rev. Sci. Instr. xv. p. 178 (1944).



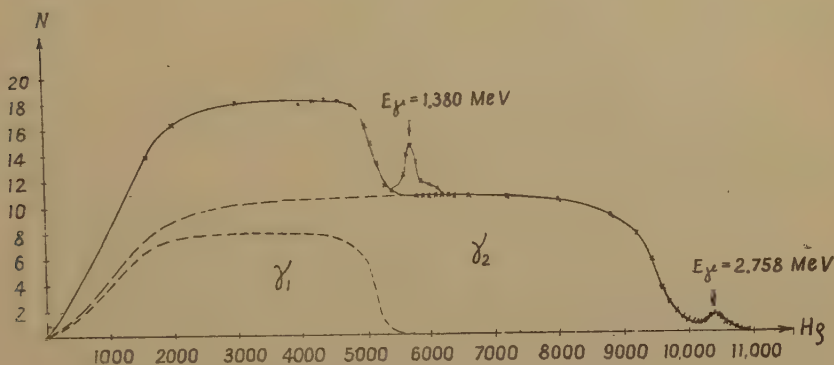
energies below this value the photo-effect in copper makes a small contribution to the secondary radiation.

A study is first made of the secondary radiation from the copper radiator. A thin lead foil (0.01–0.1 mm. thick) is then placed outside the radiator and a new energy distribution curve recorded. In this case, apart from the Compton effect from the radiator, a photo-effect from the lead foil is also obtained. When the first curve is subtracted from the second, the photo-effect in the thin lead foil results.

It proves that the photo-effect is especially marked at  $E_\gamma < 0.8$  MeV. Even at higher energies it is often possible to separate the contribution from the Compton and photo-effects. This is naturally facilitated if the demands on light-intensity can be reduced (with stronger samples), so that the spectrograph can be adjusted for higher resolution.

At higher  $\gamma$ -energies it may often be difficult, especially when the  $\gamma$ -components are closely adjacent, to distinguish the contributions of

Fig. 17.

The  $\gamma$ -radiation of  $\text{Na}^{24}$ .

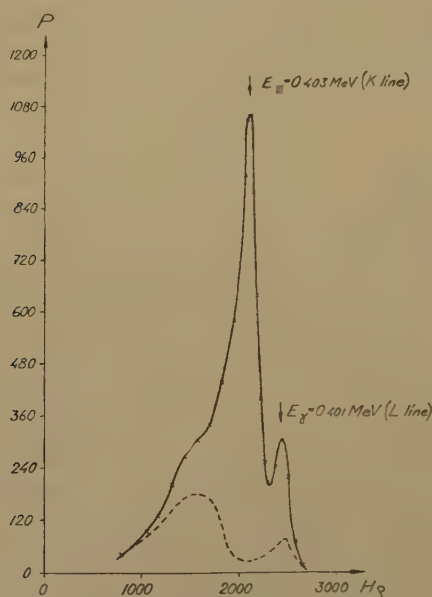
the photo-effect from the corresponding Compton edges. The photo lines are then often manifested as difficulty interpreted terracings of the high energy sides of the Compton distributions. In such cases the  $\gamma$ -energies are determined with the greatest certainty by omitting the lead foil before the radiator and thus employing the Compton effect alone.

The determination of the  $\gamma$ -energies is especially accurate when the photo-effect from the lead foil can be utilized. The high accuracy is obviously due to the fact that the peak of a well-defined line is measured. The reliability of the determination is enhanced if not only the K-photo line, but also the L-photo line of the  $\gamma$ -line in question can be measured. For a given resolving power of the spectrograph this condition is fulfilled at low energies of the  $\gamma$ -radiation, which are associated with larger distances between the K- and the L-photo lines. With a resolving power of about 3.5 per cent. it is just possible to detect the presence of the L-photo line at the side of the K-line from a  $\gamma$ -radiation of 1.37 MeV (from  $\text{Na}^{24}$ ).

Fig. 17 shows the result of an investigation by the above-described procedure of the  $\gamma$ -radiation produced in the disintegration of  $\text{Na}^{24}$  to  $\text{Mg}^{24}$ . The two long continuous energy distributions derive from the Compton effect in the copper radiator produced by two different  $\gamma$ -energies. If a lead foil 0.02 mm. thick is placed before the radiator, the two corresponding photo lines are obtained in addition to the Compton distributions. That on the left can be resolved into a K- and L-photo line.

Fig. 18 shows a recording of the  $\gamma$ -radiation from  $\text{Au}^{198}$ , generated by slow neutrons in  $\text{Au}^{197}$ . The broken curve represents the distribution found without a lead foil. At the low  $\gamma$ -energy here involved (401 keV)

Fig. 18.

The  $\gamma$ -radiation of  $\text{Au}^{198}$ .

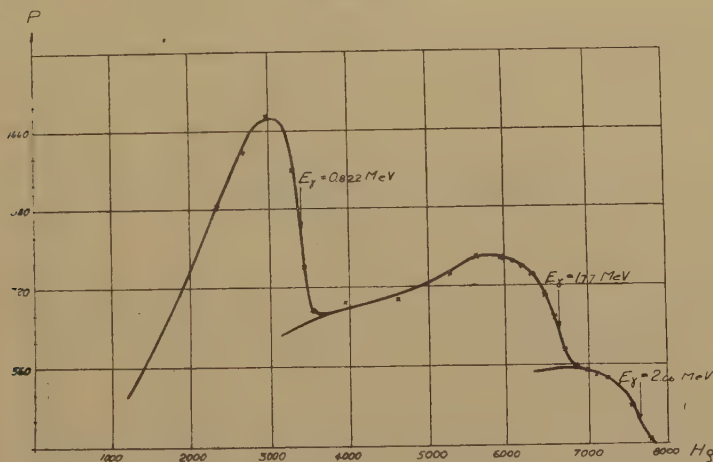
there appears, besides the Compton effect in the copper, also the corresponding K-photo line. The full curve was obtained by placing a lead foil, 0.1 mm. thick, before the radiator. The distance between the K- and L-photo lines is already so large in this case that no difficulty is experienced in separating them, notwithstanding that the spectrograph was adjusted for higher light-intensity and lower resolution (5 per cent.) than in the previous case. Fig. 18 clearly gives no less than four different possibilities for the accurate determination of the energy of the  $\gamma$ -line.

In cases where there is difficulty in utilizing the photo-effect, the  $\gamma$ -energy may be determined with moderate accuracy by measuring the inflection point at the Compton edge in the distribution curve obtained before the division of the ordinates by the corresponding abscissæ. (This division is necessary for the same reason as in the semi-circular

focusing, if it is desired to obtain a correct picture of the intensity relationships in the spectrum. Since  $\Delta(H\rho)/H\rho$  (the resolving power) is a constant of the apparatus and is independent of the energy, the width of the registered interval  $\Delta(H\rho)$  increases proportionally to  $H\rho$ . Consequently, by dividing the original curve by  $H$  we obtain the intensity for constant  $\Delta(H\rho)$  intervals.) The author has on several occasions convinced himself by empirical means that the value for the energy obtained in this manner, when inserted in the equation for the Compton effect from central impact between a photon and an electron, gives the same  $\gamma$ -energy within the limits of error as that obtained from K-photo line + the absorption energy of the K-shell (for Pb=0.088 MeV).

Fig. 19 gives a recording by the Compton method of the  $\gamma$ -radiation from the disintegration of  $\text{Mn}^{56}$  into  $\text{Fe}^{56}$  (undivided curve). The

Fig. 19.

The  $\gamma$ -radiation of  $\text{Mn}^{56}$ .

$\gamma$ -spectrum here consists of three  $\gamma$ -lines. It would be hopeless to attempt a resolution of the two highest  $\gamma$ -energies by the Wilson chamber technique or by absorption measurements. In the corresponding curve presented by Curran, Dee and Strothers (*loc. cit.*), the separation of these lines has also been unsuccessful.

As an illustration of the applicability of the technique described here for the investigation of  $\beta$ - and  $\gamma$ -spectra, it may be mentioned that it is possible, with the aid of figs. 15 and 19, to draw up the complete term scheme for the disintegration of  $\text{Mn}^{56}$ . Here both the energies of the different terms and the intensities of the different transitions between them are determined. Fig. 20 shows the appearance of this term scheme. Despite the incompleteness of the  $\beta$ -theory at present, it is possible to draw from this scheme certain conclusions, even if uncertain, concerning the states of spin and parity of the various nuclear levels.

Fig. 21 shows a quite complex  $\gamma$ -spectrum, namely, that from  $\text{ThC}$  and  $\text{ThC}''$ . A closer analysis of this spectrum shows that it can be brought

into very satisfactory agreement with the  $\alpha$ -spectrum, whereby the excited nuclear levels not partaking in the  $\alpha$ -emission can be identified

Fig. 20.

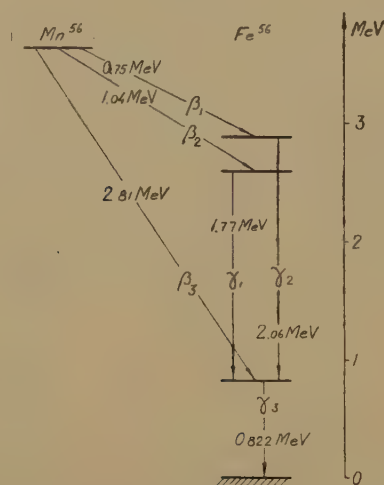
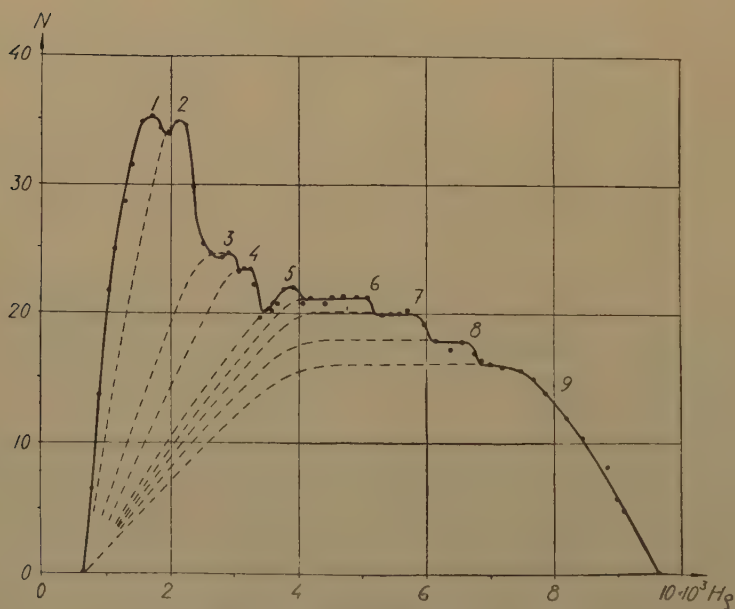
Term scheme for the disintegration of  $\text{Mn}^{56}$ .

Fig. 21.

The complex  $\gamma$ -spectrum of  $\text{Th} (\text{C}+\text{C}'')$ .

with a considerable degree of certainty\*. A remarkable fact in this connection is the essentially different appearance of this spectrum when

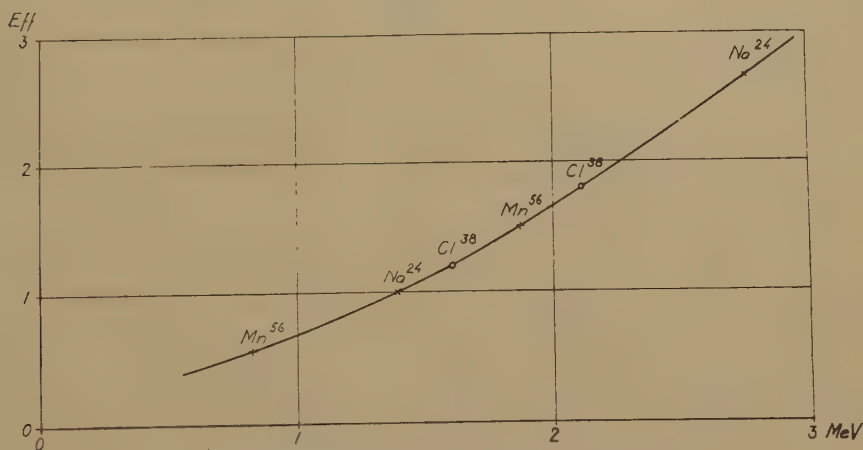
\* Cf. A. Johansson, *Ark. f. Mat., Astr. o. Fys.* (1946), in press.



it is recorded by the semi-circular method developed by Mandeville (*loc. cit.*) and Curran, Dee and Strothers (*loc. cit.*).

An important matter in the analysis of the term scheme for any radioactive disintegration is intensity relationship of the  $\gamma$ -lines. Since these magnitudes give information regarding the probabilities for transition between the various levels, it was considered justified to attempt a closer investigation as to how the relative intensities could be deduced from the experimentally determined curve for the Compton radiation. Here it is necessary, among other things, to consider the fact that both the absorption of the secondary electrons and the Compton coefficient vary with the energy of the  $\gamma$ -radiation. The area of the Compton distribution for each  $\gamma$ -energy is thus not directly proportional to the  $\gamma$ -intensity, but must be multiplied by a factor dependent on the energy, which should be derived from an efficiency curve for the radiator. Although

Fig. 22.



The efficiency curve for a copper radiator.

it is presumably possible to calculate this curve theoretically with the aid of certain simplifying assumptions, the author has chosen a more reliable empirical method\*.

In order to obtain the efficiency curve for a copper radiator, the following procedure was adopted. The areas beneath the Compton distributions for several  $\gamma$ -lines (at least two) were determined for several disintegrations, where it had been possible to determine in another manner (*e. g.*, by Fermi analysis) the term scheme and the true intensities of the  $\gamma$ -lines. By selecting the disintegration of Na<sup>24</sup> (where two  $\gamma$ -lines of different energies are emitted in cascade and therefore must be of the same intensity) and those of Mn<sup>56</sup> and Cl<sup>38</sup>, an efficiency curve was constructed which extended from 0.8 MeV to 2.8 MeV, *i.e.*, the range where the Compton method is most applicable. Fig. 22 shows the curve

\* See K. Siegbahn, "Intensities of  $\gamma$ -rays as studied by means of their Compton secondaries," Proc. Roy. Soc. A (1946), in press.

so obtained. The energies of the  $\gamma$ -radiation have been plotted as abscissæ and the efficiencies of the copper radiator as ordinates on an arbitrary scale. It follows from the curve that the efficiency of the radiator increases quite steeply with the  $\gamma$ -energy, which is naturally to be expected. Conversely, with knowledge of the areas beneath the Compton distribution from a composite  $\gamma$ -spectrum, the curve may be employed for the calculation of the relative intensities of the  $\gamma$ -lines. The curve is presumably accurate to within 10 per cent. In such an intensity determination the largest source of error lies in the resolution of the Compton distribution curve into its components, which is especially awkward at low energies.

#### GENERAL CONCLUSIONS.

With the aid of the method above described, which is based upon the use of a magnetic lens of high light-intensity and special field form, it has proved possible to make spectroscopic determinations of the energies and intensities of  $\beta$ - and  $\gamma$ -spectra when the sample available has a strength no greater than that required for a moderately good absorption measurement. A large number of radio-active disintegrations are thus made accessible for detailed study. This should be of interest in connection with the search for further general laws governing the disintegration of radio-active isotopes.

---

### XXI. *Crystal Habit and Internal Structure.*—I.

By A. F. WELLS, M.A., Ph.D.\*

[Received October 1, 1945.]

#### *Introduction.*

CLASSICAL crystallography was restricted to the study of the external forms of crystals and of their physical properties. With the development of X-ray crystallography it has become possible in many cases to determine the atomic arrangement in the interior of the crystal, and in recent years attention has been largely concentrated on this aspect of the subject. It should now be possible to begin to fuse these two aspects of crystallography into one homogeneous whole. Some progress has already been made in relating physical properties to internal structure and, conversely, a knowledge of physical properties such as optical and magnetic anisotropy is often of assistance when determining the structure of a crystal. The problem of accounting for the face development of a crystal is part of the much larger one of interatomic and intermolecular forces, for the external

---

\* Communicated by the Author.

form of a crystal is the expression not only of its internal structure but, if it is grown from solution, also of the interaction between solute and solvent.

The shapes of many crystals are sufficiently characteristic to make it possible to identify them in microchemical analysis, and many minerals are easily recognisable from their face-developments. It is not surprising, therefore, that attempts have been made to account for the shapes of crystals. There have been two lines of approach, the thermodynamical and the crystallographic. The first laid down a general condition for the stability of a crystal in equilibrium with its surroundings in terms of the surface free energies of all the faces. If the correct condition for equilibrium has been stated and if the appropriate property of the different faces could be calculated or measured, this method would obviously lead to a complete specification of the external form of the crystal. The object of the second method was to relate the face development of a crystal to its internal structure. The environment of the crystal is neglected, and only qualitative estimates of the relative stabilities of faces of different types can be made. Both these methods originated long before it was possible to determine the atomic structures of crystals. Now that we have a considerable knowledge of crystal structures it should be possible to make further progress in this field. In the present and later papers we shall review some of the earlier work, record new observations on the growth of crystals, and attempt to interpret these in terms of the structures of the crystals.

Crystals may be grown from the vapour of the substance, from the melt, or from a solution in a suitable solvent. In the first two cases we have only to consider the interactions between molecules of the substance concerned, and we should expect to find simple relations between crystal form and internal structure. The latter case is, however, more complicated, for we are concerned not only with interaction between solute molecules, but also with solute-solvent and solvent-solvent interactions. Apart from their intrinsic interest there are two reasons for dealing with these more complex cases. First, there is the practical point that most solids are purified by recrystallization from solution, and there are accordingly more data about crystals grown from solution than from the vapour or melt. Second, many interesting changes of habit may be brought about either by changing from one solvent to another or by adding a second solute to the solution.

The external form of a crystal is described in terms of (a) the types of faces present, as defined by their crystallographic indices, and (b) the relative areas of the faces of different types, if faces of more than one kind are present. Since the term "habit" is rather loosely used it will be convenient to distinguish the following types of variation in the shape of the crystals of a particular substance :—

- (1) Faces of the same *form* may be developed to different extents, *e. g.*, plates or needles instead of perfect cubes. (A *form* is a set of faces related to one another by the symmetry elements. If one

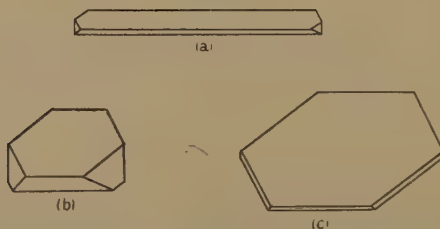
face of a particular form appears then the others should also be present, developed to the same extent on the ideal crystal.) Changes in habit of this kind are simple cases of distortion from the ideal shape due to accidental causes, *e. g.*, rapid crystallization. Such variations are usual in any batch of crystals unless they are grown carefully, hence it is usual, when illustrating a crystal, to show all faces of a given form developed to the same extent.

- (2) On a crystal exhibiting two or more forms it may be possible to vary the relative areas of the faces of different forms. This kind of change of habit is illustrated in fig. 1.
- (3) Faces of a new form appear. A simple example is the change in habit of iodoform from a hexagonal bipyramid to a prism, shown in fig. 3.

#### *Factors Influencing the Habits of Crystals.*

From the practical standpoint we may distinguish two factors which influence the shapes of crystals of a particular substance: the rate of

Fig. 1.



Crystals of naphthalene grown (a) rapidly from cyclohexane, (b) slowly from cyclohexane or ethyl alcohol, (c) rapidly from ethyl alcohol.

growth and external interactions with atoms or molecules of other kinds, *i. e.*, solvent and/or added impurities.

#### *(a) The Effect of Conditions of Growth.*

It is well known that rapid growth often leads to extreme developments. The thin plates or needles formed in rapidly cooled solutions or by sublimation are in many cases far removed from the equilibrium shape of the crystal. The simplest case is that in which the same forms are present but in different relative proportions. Naphthalene may be taken as an example. It is commonly obtained as plates  $\{001\}$  bounded by  $\{110\}$  and  $\{20\bar{1}\}$ . Rapid cooling of a solution in many solvents (*e. g.*, the lower alcohols) gives very thin plates. By slow evaporation of such a solution, however, compact crystals may be obtained. Solutions in cyclohexane give needles on rapid cooling, but again compact crystals on slow evaporation (fig. 1).

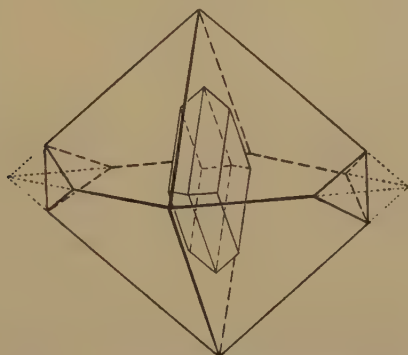
In some cases the development of extreme shapes would appear to be related to the direction(s) of the strongest interatomic or intermolecular



bonds. For example, compounds with layer structures often form thin flakes, and the thin needles of urea, potassium dihydrogen phosphate and  $\alpha$ -resorcinol which are formed by rapid cooling of supersaturated solutions represent extreme developments in the directions of the chains or spirals of hydrogen bonds. In cases such as these the different shapes are simply due to different relative developments of the same forms. In other instances faces of new forms appear which are not present on a crystal grown slowly. Two examples may be given.

*Pentaerythritol*,  $C(CH_2OH)_4$ . The equilibrium shape obtained from aqueous solution is the tetragonal bipyramid (see below). If a saturated aqueous solution is poured into acetone the crystals produced are thin square plates {001}. As will be shown later this form is very unstable in aqueous solution. Similar plates separate from a solution in  $\beta$ -ethoxyethanol on rapid cooling, whereas crystals grown slowly from this solvent are terminated by {101} bipyramid faces.

Fig. 2.



Normal habit of anthranilic acid from ethyl alcohol. Inset is a crystal tabular on {010} obtained by rapid crystallization.

*Anthranilic acid*,  $o-(NH_2)C_6H_4(COOH)$ , orthorhombic (stable modification). Crystals grown slowly from ethyl alcohol are essentially rhombic bipyramids {211}, often with small {110} (fig. 2). Rapid growth leads to diamond-shaped plates of varying thickness tabular on {010}. If a crystal of this type is replaced in saturated alcoholic solution and allowed to grow slowly it tends towards the bipyramidal shape.

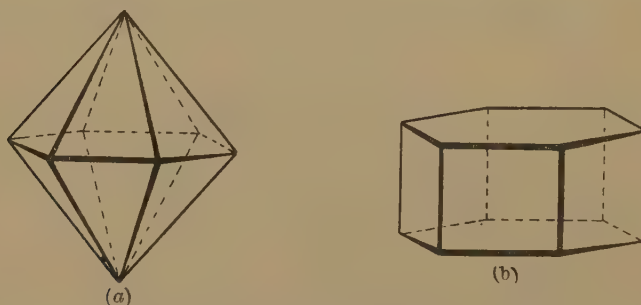
(b) *Change of Habit with Change of Solvent.*

Most of the studies of variations in crystal habit have been carried out with inorganic salts, and water has necessarily been used as solvent. By using non-ionic solutes, and particularly by choosing compounds which are soluble in a variety of solvents of chemically different types, it is possible to study the habits of crystals grown from different solvents. Many such changes of crystal habit must have been observed but comparatively few appear to have been recorded. Since they are of great importance in the present investigation, two examples are given below.

*Iodoform*,  $\text{CHI}_3$ , hexagonal. Forms observed,  $\{0001\}$ ,  $\{10\bar{1}0\}$ ,  $\{10\bar{1}1\}$ . It is normally obtained as plates  $\{0001\}$  bounded by the bipyramid  $\{10\bar{1}1\}$ . Simple bipyramids, fig. 3 (a), were obtained from a number of amines (furfurylamine, aniline, N-methyl aniline and *m*-toluidine), while the higher halogenated paraffins (*n*-butyl, octyl and dodecyl bromides), nitro-paraffins (1-nitro- and 2-nitro-propanes and 3°-nitrobutane) and cyclohexane and its derivatives (cyclohexanol, cyclohexanone and cyclohexylamine) gave crystals of type (b), fig. 3.

*Resorcinol*,  $m\text{-C}_6\text{H}_4(\text{OH})_2$ , orthorhombic. From water, compact crystals,  $\{110\}$  and  $\{011\}$ , with faces of the latter form at both ends. This development is interesting since the crystal class is *mm*. The crystals from butyl acetate show the polar development to be expected, squat prisms  $\{110\}$  terminated at one end by two  $(011)$  faces and at the other by four  $(11\bar{1})$  faces. From carbon tetrachloride and benzene, in both of

Fig. 3.



Crystals of iodoform from various solvents.

which it is very sparingly soluble, resorcinol crystallizes in very fine needles  $\{110\}$ . Further observations on the growth of crystals of resorcinol and an analysis of these results in terms of the crystal structure will be given in a later paper.

#### (c) *Changes in Habit caused by the Presence of a Second Solute.*

A considerable amount of work has been carried out on variations in crystal habit caused by the addition of "impurities" to the solution, and attempts have been made to account for these phenomena in a general way. For example, it is reasonable to assume that strong adsorption of a second solute on certain faces of a crystal will result in the development of large faces of that type on the final crystal, since the most prominent faces will be those on which deposition of material takes place most slowly (fig. 4). Following the determinations of the atomic structures of many crystals it became feasible to attempt to correlate the nature and shape of the adsorbed ions or molecules with the atomic arrangement in the different crystal faces (see, *e. g.*, Buckley (1939), France (1935, 1938, 1941), Saylor (1927). An example of a detailed study of this kind is that

by Bunn (1933) on the relation between crystal structure, adsorption-oriented overgrowth and mixed-crystal formation, with particular reference to the modification of the habits of alkali and ammonium halides by urea. In all the later work emphasis has been placed on the importance of the structures of the faces of different crystallographic types (*i. e.*, the nature and arrangement of the atoms or ions in the various faces) in relation to the structure of the adsorbed material. Little attention seems to have been devoted to the normal habit of the solute when crystallized from the particular solvent used. It would appear more logical first to attempt to account for this before considering changes in habit induced by adding foreign substances.

It will be apparent that the study of crystal habit is closely related to the study of solubility, for the crystal face is the dividing surface between the crystalline solid and saturated solution. For a solute A in a solvent B the crystal structure of the solute is determined by the interactions A-A, the crystal habit by A-A and A-B, and the solubility by A-B (and also B-B and A-A). Therefore, we shall adopt the view-point that all such

Fig. 4.



Rapid growth normal to face A leads to disappearance of that face.

phenomena as the dissolution of a crystal, the separation of a solid phase as a colloidal suspension, as compact crystals or as fine needles, must be regarded as manifestations not only of the way in which the primary particles of the solute interact one with another but also of the way in which solute and solvent interact. The very fact that many substances crystallize with different habits from different solvents shows that we must generalize the idea of habit modification by adsorbed impurities and extend it to include interaction with the solvent. If this is very different for different faces it will lead to habit modification just as in the case of the preferential adsorption of impurities. These considerations indicate the restricted validity of any relations between internal structure and crystalline form which do not take account of the solvent. We shall show that the effect of the solvent on crystal habit ranges from one extreme to the other. For ionic crystals (*e. g.*, alkali halides) crystallizing from water, interaction with the solvent is of negligible importance in determining habit, except possibly for the fluorides. Accordingly the general face development is predictable from the crystal structure. On

the other hand, the nature of the solvent has a considerable effect in the case of, for example, some hydroxy-compounds, as already mentioned. This is not surprising when we remember that a compound like resorcinol is insoluble in benzene and cyclohexane but extremely soluble in water, cyclohexanol and cyclohexanone. There is a correspondingly large difference between the habits of crystals of resorcinol grown from water and benzene.

TABLE I.

Point-groups in which the general or a single special form is a polyhedron.

Point-group Symbol	General Form ( <i>hkl</i> )	Special Forms
222	bisphenoid	—
<i>mmm</i>	bipyramid	—
4	bisphenoid	bisphenoid <i>hhl</i> or <i>h0l</i>
$\bar{4}2m$	scalenohedron	{ bisphenoid <i>hhl</i> bipyramid <i>h0l</i>
42	trapezohedron	bipyramid <i>hhl</i> or <i>h0l</i>
$4/m$	bipyramid	bipyramid <i>hhl</i> or <i>h0l</i>
$4/mmm$	{ ditetragonal } bipyramid	bipyramid <i>hhl</i> or <i>h0l</i>
$\bar{3}$	rhombohedron	rhombohedron <i>hhl</i> or <i>h0l</i>
32	{ trigonal trapezohedron }	{ bipyramid <i>hhl</i> rhombohedron <i>h0l</i>
$3/m$	bipyramid	bipyramid <i>hhl</i> or <i>h0l</i>
$\bar{3}m$	{ ditrigonal scalenohedron }	{ hexag. bipyramid <i>hhl</i> rhombohedron <i>h0l</i>
$3/mmm$	{ ditrigonal bipyramid }	{ hexag. bipyramid <i>hhl</i> trig. bipyramid <i>h0l</i>
62	trapezohedron	bipyramid <i>hhl</i> or <i>h0l</i>
$6/m$	{ hexagonal bipyramid }	bipyramid <i>hhl</i> or <i>h0l</i>
$6/mmm$	{ dihexagonal bipyramid }	bipyramid <i>hhl</i> or <i>h0l</i>
and the five cubic classes, 23, 432, <i>m</i> 3, $\bar{4}3m$ , and <i>m</i> 3 <i>m</i>		

### *The Equilibrium Shape of a Crystal.*

A crystal is a solid body bounded by a number of plane surfaces forming a closed convex polyhedron. Of the 32 crystal classes 20 may be distinguished from the rest on the grounds that the general form or a special form is a polyhedron, *i. e.*, the set of faces generated by the symmetry elements forms a closed polyhedron. In the other classes faces of at least two forms are required to give a polyhedron. Of course, even in the former (20) classes the general form is not necessarily developed, and the crystal may exhibit two or more forms. In Table I.



are listed those forms which enclose a polyhedron. The cubic classes are omitted since in these all forms are of this type. Judging by the published data, crystals (other than cubic) showing only one form are rare.

We may therefore state our problem as follows :—

(1) If the faces of a crystal are all of one form, we require to know why this and not some other form develops, *e. g.*, why NaCl grows into cubes rather than octahedra or dodecahedra, all of which are consistent with the symmetry of the internal structure of the crystal. In the case of crystals of less than holohedral symmetry it will be necessary also to account for the smaller number of faces in certain forms. Taking urea as an example, we require to show why the form  $\{111\}$  in the class  $\bar{4}2m$  consists of only four faces, *i. e.*, to show in terms of the atomic structures of the faces why the face (111) is more stable than  $(\bar{1}\bar{1}\bar{1})$ , which does not occur.

(2) If faces of more than one form are present we have to show, first, why these particular forms develop and second, what determines the relative areas of the faces of different types.

It is well known that crystals of a given substance differ as regards the relative developments of faces of different forms. Before proceeding further it is essential, therefore, to know whether there is an equilibrium shape for a crystal of a particular substance grown from a certain solvent.

### *The Concept of Minimum Total Surface Free Energy.*

The condition for stability of an isolated drop of a fluid is that its surface free energy, and hence its area, is a minimum. Gibbs suggested that a similar condition should apply to a crystal, *viz.*, that for a crystal in equilibrium with its surroundings at constant temperature and pressure, its total (Gibbs) free energy is a minimum for a given volume. (Curie, 1885, arrived at the same result by a different method.) If we suppose that the volume free energy per unit volume is constant throughout the crystal, then the condition becomes  $\sum_1^N A_i g_i = \text{minimum}$ , where  $g_i$  is the surface free energy per unit area of the  $i$ th face of area  $A_i$  on a crystal bounded by  $N$  faces. In other words, those faces develop which lead to a minimum total surface free energy for a given volume, so that any crystal, if kept for a sufficiently long time in a saturated solution, should attain the equilibrium shape determined by this criterion. If the above condition applies to crystals of all sizes it follows that (1) two or more geometrically similar crystals of different sizes (each equilibrated in itself) brought into the same saturated solution should ultimately become one crystal, and (2) the solubility of a crystalline material is a function of the size of the crystals. We shall shortly consider these points, but first we should point out that Gibbs realized that, owing to the different degrees of internal order in crystals and liquids, the practical realization of an equilibrated crystal might be difficult. In a drop of liquid there is random arrangement of the molecules (or atoms) so that every portion of the surface is similar in structure to every other. In a crystal, however, the structural units are arranged in a regular way in three dimensions so that

the whole array possesses certain symmetry. The bounding surfaces of the crystal are plane but not necessarily equivalent, and there is an infinite series of sets of possible faces for any crystal. Gibbs appreciated that the growth of a crystal is a periodic process, in contrast to one such as condensation or evaporation, for it takes place by the deposition of new layers of material on the faces of the crystal. There might, therefore, be "activation energies" associated with the commencement of a new layer, and these might be independent of the area of the face. He concluded that "it seems not improbable that the form of very minute crystals in equilibrium with solvents is principally determined by the condition that  $\Sigma_1^N(A_i g_i)$  shall be a minimum for the volume of the crystal, but as they grow larger (in a solvent no more supersaturated than is necessary to make them grow at all), the deposition of new matter on the different surfaces will be determined more by the nature (orientation) of the surfaces and less by their size and relations to the surrounding surfaces. As a final result, a large crystal thus formed will generally be bounded by those surfaces alone on which the deposition of new matter takes place least readily, with small, perhaps insensible, truncations. If one kind of surface satisfying this condition cannot form a closed figure, the crystal will be bounded by two or three kinds of surfaces determined by the same condition. The kinds of surface thus determined will probably generally be those for which  $g_i$  has the least values. But the relative development of the different kinds of sides will not be such as to make  $\Sigma_1^N(A_i g_i)$  a minimum. The growth of the crystal will finally be confined to sides of a single kind."

Wulff (1901) showed that the Gibbs equilibrium shape of a crystal is very simply related to the relative surface free energies of the faces. The crystal should form a polyhedron such that the perpendicular distances from a point within the crystal to the faces are proportional to the surface free energies of the appropriate faces. von Laue (1943) has recently reviewed, and added to, the various proofs of this theorem of Wulff, and emphasizes that it is not sufficient merely to determine, for any arbitrary combination of faces, the shape corresponding to minimum surface free energy. It would be necessary to consider all possible combinations of faces and to find which of the relative minima is actually the smallest. Now the validity of the minimum surface free energy criterion could only be tested directly if the relative values of these energies of sufficient (ideally, of all the possible) faces of a crystal were known, and unfortunately very little is known of these quantities. Indirect evidence has, however, been adduced to show that the Gibbs condition is not normally applicable to crystals of radius greater than about  $10^{-4}$  cm. The Thomson equation giving the ratio of the vapour pressure of a drop of liquid to that of a plane surface in terms of the surface tension of the liquid and the radius of the drop shows that for a drop with radius  $10^{-4}$  cm. the vapour pressure is only 0.1 per cent. greater than that of a plane surface. The solubility of crystals should vary with their radius and interfacial tension

solid-liquid ( $\sigma$ ) according to the equation

$$C_r = C_0 e^{\frac{2\sigma M}{RT\rho r}},$$

where  $C_r$  is the solubility of a crystal of radius  $r$ ,  $C_0$  the concentration of a solution which would be in equilibrium with an indefinitely large crystal,  $M$  the molecular weight and  $\rho$  the density of the solute. Hulett (1901) found, by conductivity measurements, that the solubility of gypsum particles of radius  $0.3 \mu$  was 18.2 millimoles/litre at  $25^\circ$ , whereas that of crystals with radius  $2 \mu$  was the normal value, 15.3 millimoles/litre. Although the precise meaning to be attached to such "solubilities" is doubtful, particularly if the crystals have been ground up (see, for example, the discussion by Valetton (1924)), it has been generally concluded that the differences between the surface free energies of different faces are not the only factors operative in determining the shapes of macroscopic crystals (see also Stranski and Kaischew (1934, 1935, 1937), Stranski (1943) and Volmer (1939)).

Our first task is clearly to ascertain whether there is in fact an equilibrium shape for macroscopic crystals of a particular substance grown under specified conditions. Only if this is so will it be possible ultimately to relate the morphology of a crystal to its internal structure and its interactions with the solvent. Evidence of three kinds may be sought. First, large numbers of crystals may be grown slowly to see if they all develop in the same way. Second, crystals of other shapes obtained, for example, by rapid crystallization, may be grown larger under more "ideal" conditions to see if they approach a standard shape. A third type of experiment is possible in cases where crystals grown from different solvents show different habits. Evidence of all three kinds will be presented later. The interpretation of the results of experiments of the second type raises a number of questions.

### *The Practical Realization of Equilibrium Shape.*

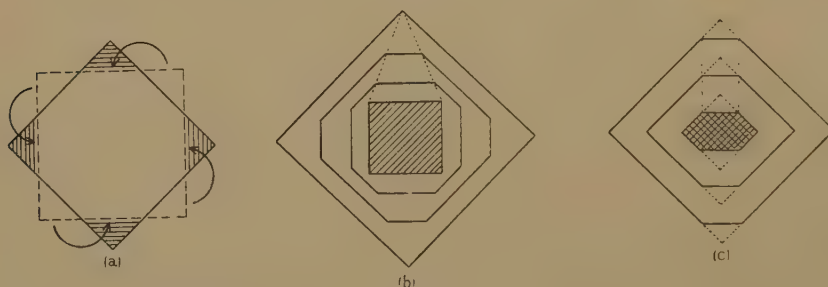
As already mentioned, Gibbs foresaw that there might be difficulty in commencing a new layer on a particular face of a crystal. There is, however, a further complication. Let us suppose that there is a definite equilibrium shape for a crystal growing from a certain solution and that a crystal of non-equilibrium shape is placed in the saturated solution, which is not allowed to evaporate. Material should be transferred, at some unspecified rate, from certain parts of the crystal via the solution to other parts until the equilibrium shape is reached. We may imagine that the original crystal has the shape indicated by the broken lines in fig. 5 (a), which might represent a section of a tetragonal prism {100}, and that the equilibrium shape is that enclosed by full lines (the prism {110}). Material would have to be transferred as shown by the arrows. The initiation of the process would consist in the movement of molecules from one part of a (100) face to another part of the same face, and moreover, there would have to be a nett overall transfer of molecules to the



centre of the face. We should expect that the probability of this process ever taking place would be very small. If, on the other hand, evaporation of the solution is permitted, molecules may be added on all faces of the crystal, which can then grow as shown at (b).

Failure to observe any tendency towards an equilibrium shape has already been recorded (*e.g.*, Herzfeld, 1930). We have found that crystals of pentaerythritol of various shapes show no change after shaking for many days in a sealed vessel containing saturated solution at 25° C., though very different results are observed if evaporation is allowed (see below). In view of what we have said it is evident that the failure to change shape spontaneously is not to be interpreted to mean that there is no equilibrium shape. A further point now arises which is of importance in interpreting the growth of crystals of non-equilibrium shape. Is it necessarily possible for a crystal of *any* initial shape to reach the equilibrium shape? This obviously depends on the relative rates of growth on the various faces. If, for example, the crystal in fig. 5 (c)

Fig. 5.



The growth of crystals towards equilibrium shape.

grows as shown, it can never reach the equilibrium shape, though it will approach it.

The following new data on the growth of crystals of certain organic compounds can now be presented. We may summarize under three heads the evidence at present available for the existence of an equilibrium shape for crystals of a particular substance grown from a given solvent :

- (i) Constancy of shape of crystals grown from one solvent.
- (ii) Growth of crystals of other shapes towards the equilibrium shape.
- (iii) Change in equilibrium shape with change of solvent.

(i) *Constancy of Shape of Crystals Grown from one Solvent.*

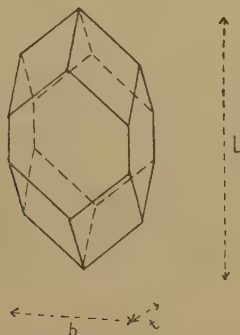
Measurements have been made of the dimensions of crystals of anthranilic acid grown from aqueous solution to determine whether in fact the same shape is maintained during growth. Large numbers of crystals were grown and batches of ten were removed and measured at intervals. Measurements were made of the overall length and of the



breadth and thickness (see fig. 6), and the mean values of the ratios length/breadth and thickness/breadth were calculated.

Owing presumably to the fact that the crystals were grown resting on a (100) face there is some drift in these values (in particular the crystals grew thinner) which cover a volume range of nearly 200 : 1. More constant values for these ratios would be obtained from experiments with

Fig. 6



Crystal of anthranilic acid from aqueous solution

crystals suspended in the solution, as we shall show in a later paper, but these rough results are sufficient to show that there is no appreciable alteration in the general shape of these crystals during growth to some 200 times their original volume.

TABLE II.

Dimensions of Crystals of Anthranilic Acid (in mm.)

	L.	b.	t.	L/b.	t/b.
Set 1	1.17	0.74	0.31	1.58	0.41
„ 2	2.47	1.39	0.55	1.78	0.39
„ 3	8.50	4.66	1.49	1.82	0.32

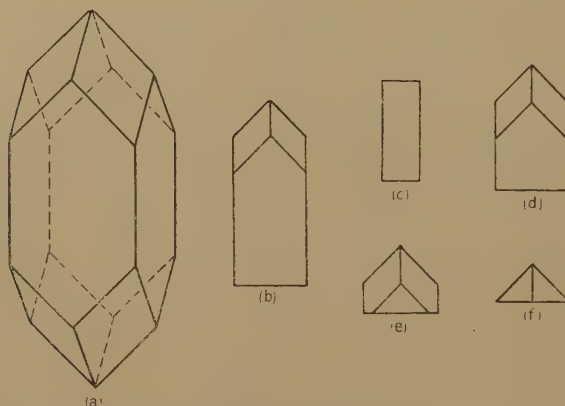
(ii) *Growth of Crystals of Other Shapes towards Equilibrium Shape.*

Evidence on this point has already been presented for crystals of anthranilic acid growing in ethyl alcohol solution. A more detailed study has been made of pentaerythritol,  $C(CH_2OH)_4$ .

Crystals of type (a), fig. 7, may be readily grown from aqueous solution. Referred to the tetragonal cell with  $a=6.10$  Å and  $c=8.73$  Å, the forms are  $\{110\}$ ,  $\{101\}$  and  $\{10\bar{1}\}$ . The crystals may be twins, since the point-symmetry is only  $\bar{4}$ , but this is immaterial in the present study. Crystals terminated at one end by (001), shown in elevation at (b), may be

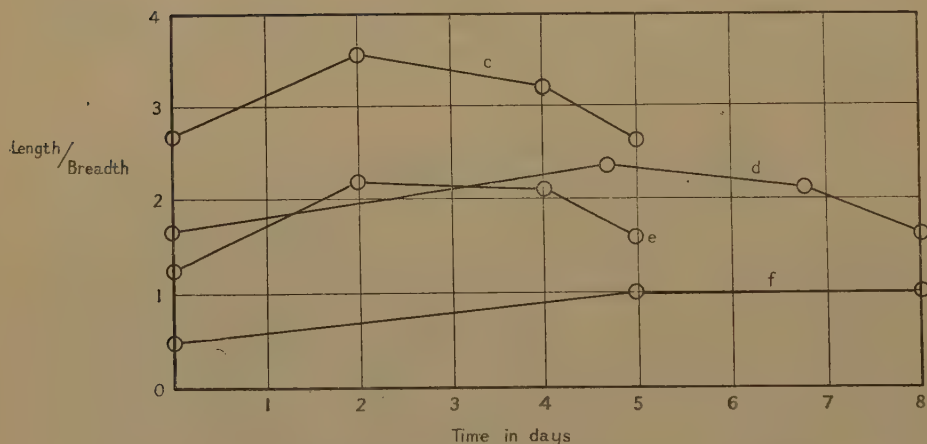
obtained by cooling a hot saturated solution and allowing crystals to grow downwards from the surface. Alternatively a crystal of type (a) may be cleaved, there being a perfect (001) cleavage. Crystals of type (b) kept in a saturated solution at constant temperature showed no change in shape after several days when evaporation was prevented, but if placed

Fig. 7.



Crystals of pentaerythritol.

Fig. 8.



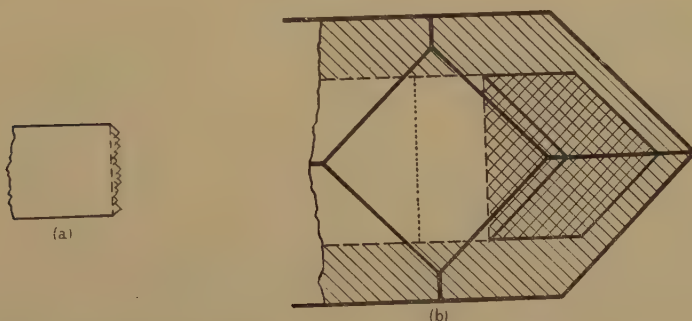
Pentaerythritol growth curves.

in saturated solution in an open beaker, rapid growth took place at one end until, after 2-3 hours, the (001) face was completely replaced by bipyramid faces. This is described in more detail below. The experiment was repeated with a series of crystals (c)-(f), and measurements of length and breadth were made at intervals. The results are shown in fig. 8. The initial rise in the curves is due to the completion of the ends of the crystals, which then all tend towards a definite limiting shape. This is the

bipyramid, though small prism faces were still present on many crystals when the experiment had to be concluded owing to the deterioration of the crystals.

A more detailed study of the growth of a crystal of type (b) was made under the microscope. Deposition of new material on (001) did not take place in layers, but instead a large number of minute pyramids developed, *i. e.*, the (001) face was replaced as quickly as possible by (101) faces (fig. 9, a). Growth then continued as shown at (b). Only after the pyramid faces had been completed on the end of the crystal did any appreciable deposition take place on the (110) faces. The crystal then grew in such a way as to reduce the area of (110) faces relative to (101), the crystal tending towards a bipyramid. Fig. 9 (b) is drawn to scale from the actual measurements made under the microscope. A cleavage line (dotted) served as a base-line for measurements along the *c*-axis. These observations show that the (001) face is highly unstable in aqueous solution, and that the relative stabilities of the faces are  $(101) > (110) > (001)$ . These facts will be discussed in relation to the internal structure of the crystal in a later paper.

Fig. 9.



Growth of a crystal of pentaerythritol. In (b) the unshaded portion represents the original crystal, and the two kinds of shading indicate two stages in the growth of the crystal.

### (iii) *Change in Equilibrium Shape with Change of Solvent.*

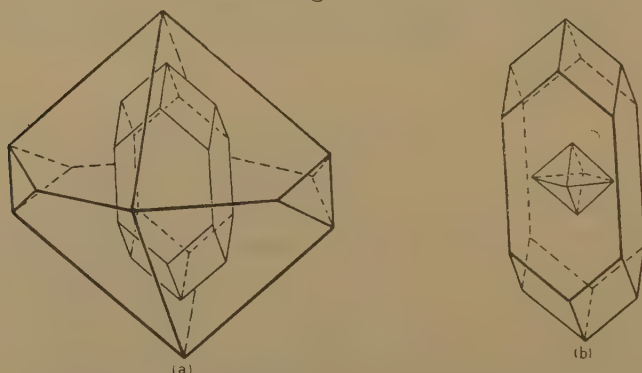
Perhaps the most conclusive proof that there is a definite equilibrium shape for a crystal grown from a given solvent is provided by experiments of the following type. A crystal of the substance is grown slowly from solvent I and then transferred to a saturated solution in solvent II from which the substance crystallizes with a markedly different habit. The crystal continues to grow but now develops the habit corresponding to solvent II. Anthranilic acid is a convenient solute for illustrating this point. A crystal grown from aqueous solution, tabular on (100), was suspended in ethyl alcohol solution. Growth took place as shown in fig. 10 (a), the final crystal being essentially a bipyramid. Crystals of anthranilic acid from glacial acetic acid are also prismatic like those from water, though with rather different relative dimensions. If, therefore, a

crystal grown from alcoholic solution (bipyramid) is regrown in acetic acid solution, it develops into a prismatic crystal as shown in fig. 10 (b).

### Conclusion.

We have seen that a number of factors can influence the habit of crystals of a (given polymorphic form of a) substance—rate of growth, nature of the solvent, and the presence of impurities. However, all the crystals from a solution in a given solvent, if grown sufficiently slowly, do exhibit a standard shape which is characteristic of the particular solute-solvent combination. The influences tending to restrain the crystals to the equilibrium shape will naturally vary greatly according to the nature of the solute and solvent, and it is to be expected that greater divergences from the equilibrium shape will be observed, and will result from smaller variations in the conditions of crystallization, in some systems than in others. It has been shown that crystals of non-equilibrium shape do

Fig. 10.



(a) Crystal of anthranilic acid from aqueous solution (inset) grown larger in ethyl alcohol solution. (b) Crystal from alcohol (inset) grown larger in acetic acid solution.

tend, on further growth, towards the standard shape. Such changes do not, however, take place unless growth is permitted. This fact, together with indirect evidence from the dependence of solubility on particle size, indicates that the criterion of minimum surface free energy, originally postulated by Gibbs as the condition for stability of a crystal, cannot be applied in its original simple form to a macroscopic crystal which has been grown at a finite rate. An essential difference between a drop of fluid and a crystal is that the former can adjust its shape by the rearrangement, within itself, of its molecules. This is not generally possible at ordinary temperatures for a crystal. Thus a crystal of non-equilibrium shape in a saturated solution could rearrange itself to the equilibrium shape only by the passage of molecules *via the solution* from one face on to others or on to other parts of the same face. Such processes do not take place at measurable rates in the cases of macroscopic crystals, except possibly at temperatures near the melting point. The fact that the final shape of large crystals is determined by the relative rates of



growth in different directions, a conclusion reached by Gibbs, Valetton and others, would seem indisputable. The problem of accounting for the equilibrium shape therefore consists in discovering what factors determine (1) the choice of planes on which growth takes place, and (2) the relative rates of growth on those planes.

Consider first a crystal built of atoms of one kind. An atom in a face of the crystal has fewer nearest neighbours than one in the interior, and this asymmetry of environment is different for faces of different crystallographic types. We might therefore expect to be able in such cases to relate the nature of the surface to its stability, and hence to the rate of deposition on that surface. If, however, there are atoms of more than one kind in the crystal, faces of different types may differ in another respect, viz., the nature of the atoms in the faces. In such a case stronger interaction between one atomic species and the solvent may well be the decisive factor in determining the relative rates of deposition of solute molecules on the various faces. In Part II. we shall examine the validity of relationships which are supposed to exist between the face developments and internal structures of crystals.

#### Summary.

Some of the factors influencing the habits of crystals are discussed. In many cases interaction with the solvent is important. The Gibbs condition for the stability of a crystal, viz., that for a given volume the total surface free energy shall be a minimum, is discussed. Evidence of various kinds is adduced to show that there is a definite equilibrium shape for crystals of a given solute crystallizing from a particular solvent, but it is also shown that a macroscopic crystal of non-equilibrium shape shows no tendency to approach this equilibrium shape unless growth is permitted. The equilibrium shape is a function not only of the internal structure of the crystal, but also of the interactions between the atoms (molecules) in the various faces of the crystal with the solvent and other molecules in the solution.

#### References.

- Buckley, Proc. Manchester Lit. Phil. Soc. lxxxiii. p. 31 (1939).  
 Curie, *Bull. Soc. min. de France*, viii. p. 145 (1885).  
 France, J. Phys. Chem. xxxix. p. 425 (1938), xlii. p. 1079 (1938), xlv. p. 395 (1941).  
 Gibbs, 'Collected Works,' i. p. 320.  
 Herzfeld, Colloid Symposium, vii. p. 51 (1930).  
 Hulett, *Z. Phys. Chem.* xxxvii. p. 385 (1901).  
 Laue, *Z. Krist.* cv. p. 124 (1943).  
 Saylor, Colloid Symposium, v. p. 192 (1927).  
 Stranski, *Z. Krist.* cv. p. 90 (1943).  
 Stranski and Kaischew, *Z. phys. Chem.* B, xxvi. pp. 81, 114, 312 (1934), *ibid.* B, xxxv. p. 427 (1937); *Physik. Z.* xxxvi. p. 393 (1935).  
 Valetton, *Z. Krist.* lx. p. 31 (1924).  
 Volmer, 'Kinetik der Phasenbildung,' p. 91 (Dresden, 1939).  
 Wulff, *Z. Krist.* xxxiv. p. 449 (1901).

Imperial Chemical Industries, Ltd.,  
 Research Laboratories,  
 Blackley, Manchester 9.

XXII. *On a Difficulty in the Application of the Law of Mass Action.*

By J. A. CHALMERS, M.A., Ph.D., and C. W. GIBBY, M.Sc., Ph.D.\*,  
Durham Colleges in the University of Durham.

[Received March 11, 1946.]

THE purpose of the present note is to draw attention to a difficulty sometimes encountered by students in the numerical application of the law of mass action and its solution, for it is not adequately dealt with in any text-book of physical chemistry with which we are familiar.

In the derivation of the law of mass action from kinetic principles for a reaction such as  $A+B=C+D$ , the equilibrium constant is obtained by equating the velocities of reaction in opposite directions. If the chance of collision between one specified molecule of A and one specified molecule of B is  $n$  per unit time, the total number of collisions between molecules of A and molecules of B in unit time is  $n[A][B]$ , square brackets denoting concentrations. If a fraction  $\alpha$  of these collisions results in reaction, the number of reactions is  $n\alpha[A][B]$ . Similarly, the number of reactions in the reverse direction in the same time may be written as:  $n'\beta[C][D]$ . At equilibrium these two numbers will be equal, and so

$$\frac{[C][D]}{[A][B]} = \frac{n\alpha}{n'\beta} = K.$$

It would seem possible to apply this result to a reaction such as  $H_2+I_2=2HI$  by writing

$$\frac{[HI]^2}{[H_2][I_2]} = K.$$

If the reaction is between  $x$  molecules of  $H_2$  and  $y$  molecules of  $I_2$  per unit volume at the beginning and results in the production of  $2z$  molecules of HI, the total volume remaining constant, the equilibrium constant has the value

$$K=4z^2/(x-z)(y-z).$$

If the reaction be written as  $H_2+I_2=HI+HI$ , by analogy with  $A+B=C+D$ , then the constant becomes, instead,

$$K=z^2/(x-z)(y-z).$$

This, however, has assumed that the decomposition of HI takes place only when a collision occurs between a member of one group of HI molecules and a member of the other group; for clarity the right-hand side of the equation may be re-written as  $HI+HJ$ .

Decomposition occurs (a) by collision between HI and HI, (b) by collision between HI and HJ, or (c) by collision between HJ and HJ.

---

\* Communicated by the Authors.

The concentrations of HI and HJ are each  $z$ , but the number of collisions in classes (a) and (c) will be  $\frac{1}{2}n'z^2$  while that in class (b) will be  $n'z^2$ . The reason for this difference is easily seen, but the fact usually appears to escape notice. There are  $z$  ways of choosing one HI (as distinct from HJ) molecule to participate in a collision and  $(z-1)$  (which is equal to  $z$  when  $z$  is large) ways of choosing the second HI molecule; there are not  $z^2$  ways of choosing the pair of molecules for the collision, but only  $\frac{1}{2}z^2$ , since in the above argument every pair of molecules is reckoned twice. For collisions between HI and HJ there is not the same duplication and so the number of collisions is  $n'z^2$ . Adding all the collisions, the total is  $2n'z^2$  and so

$$K = \frac{n\alpha}{n'\beta} = 2z^2/(x-z)(y-z).$$

Considering the HI molecules as all identical, the number of collisions is  $\frac{1}{2}n'(2z)^2 = 2n'z^2$  as before, and the apparent contradiction has disappeared.

The confusion which sometimes arises is due to the use of statements that the velocity of decomposition of HI is proportional to the square of the concentration of HI, or to the product of the concentrations of HI and HJ. These statements are both true, but the factors of proportionality are different in the two cases.

If the actual value of  $K$  is to be used to obtain information about  $n$ ,  $n'$ ,  $\alpha$  or  $\beta$ , then it must be put in the form

$$K = \frac{n\alpha}{n'\beta} = 2z^2/(x-z)(x-y),$$

and not in either of the forms  $K = \frac{[\text{HI}]^2}{[\text{H}_2][\text{I}_2]}$  or  $K = \frac{[\text{HI}][\text{HJ}]}{[\text{H}_2][\text{I}_2]}$ .

### XXIII. *The Viscosities and Structures of Acid Chlorides and Amides.*

By J. NEWTON FRIEND and WILLIAM D. HARGREAVES\*.

[Received January 8, 1946.]

THE viscosities of several acid chlorides and amides have been measured at various temperatures from atmospheric up to as near the boiling point as practicable; the boiling-point viscosities have been calculated with the aid of Sheppard's equation<sup>(1)</sup>,

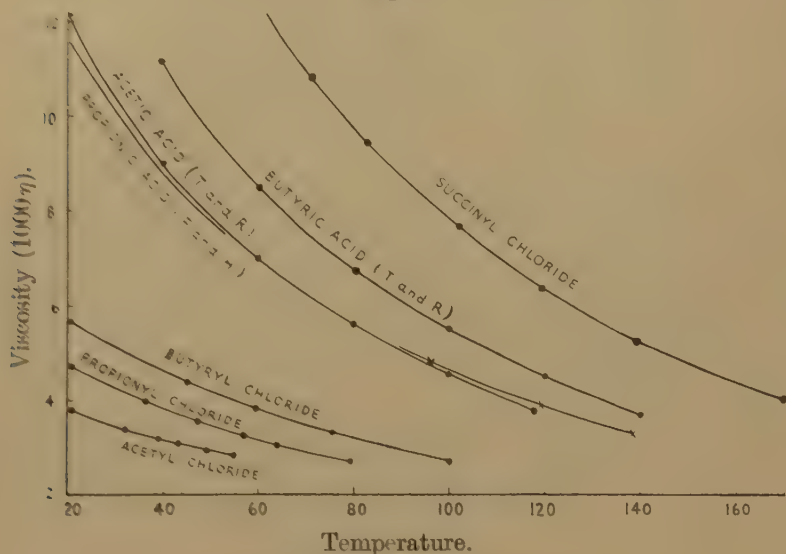
$$\log 10^3\eta = A + B/T.$$

The viscosity temperature curves of the acid chlorides are shown in fig. 1. They rise steadily with the molecular weight. Those for the corresponding free acids are considerably higher, due to association, and do not show a regular increase. It has already been shown that the

\* Communicated by the Author.

rheochor<sup>(2)</sup> is an additive atomic and constitutive function of the molecule: the following boiling-point values are used below: C 12.8, H 5.5, O 10.0, Cl 27.3, N (when negative) 6.6, ketonic double bond 3.2. A contraction occurs on 5-membered ring formation which varies slightly with the nature of the ring; it is approximately one unit.

Fig. 1.



T and R—Thorpe and Rodgers<sup>(14)</sup>.  
F and H—Friend and Hargreaves<sup>(2)</sup>.

The observed and calculated rheochors of the chlorides and free acids are as follows:—

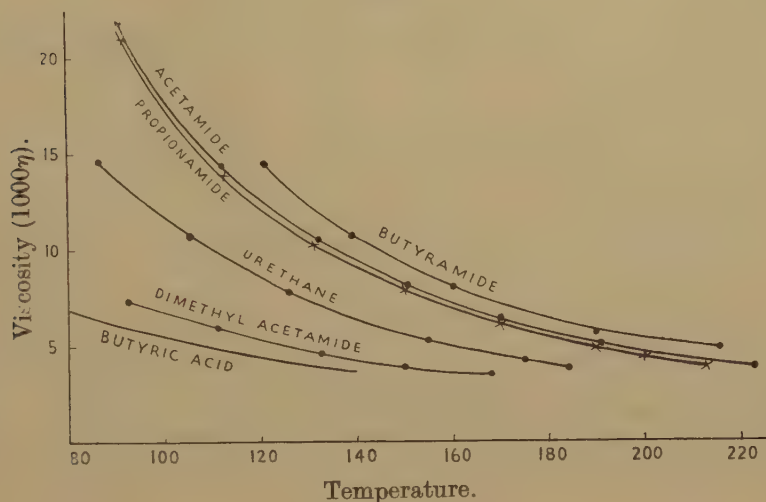
	R (obs.).	R (calc.).		R (obs.).	R (calc.).
Acetyl chloride ..	84.1	82.6	Acetic acid ..	75.4	70.8
Propionyl „ ..	106.6	106.4	Propionic „ ..	99.1	94.6
Butyryl „ ..	129.9	129.2	Butyric „ ..	123.5	118.4
Succinyl „ ..	154.0	154.2 (symm.) 150.0 (cyclic)			

The observed rheochor of acetyl chloride is 1.5 units too high; this is usual for initial members of series. Gibling<sup>(3)</sup> has observed analogous irregularities in the parachor, and points out that an end-carbon does not contribute quite the same to the molar volume as the other carbon atoms. The agreement between the observed and calculated values for propionyl and butyryl chlorides shows that these substances are monomeric: the corresponding acids, however, are known to be associated, hence their higher viscosities. The data for succinyl chloride show that it is essentially the symmetrical form. A study of the parachor led to the same conclusion<sup>(4)</sup>.



The viscosity-temperature curves for the amides are shown in fig. 2. Those for acetamide and propionamide lie close together and are lower than for *n*-butyramide. The curves for the corresponding acids (fig. 1) behave similarly. All are associated through their hydrogen bonds and their observed rheochors are high. With *N*-dimethyl acetamide, however, such association is not possible; its viscosity is thus much less than that of its isomeride, *n*-butyramide, and appreciably less also than the viscosities of propionamide and acetamide, despite their lower molecular weights. Its rheochor is normal, which confirms its monomeric character.

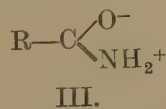
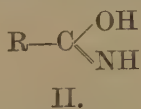
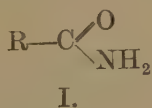
Fig. 2.



	$R_{obs.}$	$R_{calc.}$	$B/T_b.$
Acetamide .....	79.8	72.9	1.86
Propionamide .....	103.3	96.7	2.01
<i>n</i> -Butyramide .....	127.1	120.5	2.39
Urethane .....	113.4	106.7	2.13
<i>N</i> -Dimethyl acetamide .....	120.6	120.5	1.52

The relatively high values of  $B/T_b$  for the first four amides again indicate appreciable association, for  $B$  (Sheppard's equation)  $\propto$  molar latent heat at the boiling point<sup>(5)</sup>.

There are three possible structures for amides, namely the usually accepted normal I., the iso-amide II. and the betaine III. :—



The rheochor enables us to distinguish between I. and II. Both H-atoms of an  $\text{NH}_2$  group can yield bonds of the type  $\text{N}-\text{H} \dots \text{N}$ , each of which

raises the rheochor by an average of 1.5 units. The viscosities of normal amides should thus be of the same order as those of amines of analogous molecular weight. Actually they are much greater as is seen by a comparison of *n*-butylamine and butyramide<sup>(7)</sup>. Their viscosities exceed even those of the corresponding free acids (fig. 2). This suggests the iso-structure which contains, in addition to one N—H...N bond, one O—H...O, which raises the rheochor by 4.5 units. The third structure would also contribute to high viscosity just as  $\text{—N}^+\text{O}^-$  has been shown to do<sup>(2)</sup>, but at present data are insufficient to enable its rheochor to be calculated.

Assuming the effect of the double bond between C and N to be the same as between C and O (it must be of the same order), the following data are obtained :—

	R (obs.).	R (calc.).	R (calc.).
		iso.	normal.
Acetamide .....	79.8	78.9	75.9
Propionamide .....	103.3	102.7	99.7
<i>n</i> -Butyramide .....	127.1	126.5	123.5
Urethane .....	113.4	112.7	109.7

At their boiling points, therefore, the amides do not possess to any appreciable extent the normal structure. This is in agreement with Hantsch's<sup>(8)</sup> interpretation of the U.V. absorption spectra of certain other amides.

#### *Experimental.*

The acetyl and butyryl chlorides were supplied by B.D.H. and were fractionated in a thoroughly dried apparatus under a pressure of 46 cm. The propionyl chloride was synthesized from re-distilled acid and phosphorus trichloride and fractionated. Succinyl chloride was obtained by interaction of the "Analar" acid with phosphorus pentachloride, fractionated several times under reduced pressure until the first and last fractions both melted sharply at 16.7° C.

Acetamide was recrystallized twice from a mixture of benzene and dry ethyl acetate. Urethane was twice distilled, the fraction used having a constant boiling point. Propionamide and *n*-butyramide were prepared by refluxing the corresponding ammonium salts with the acid so that only water escaped, and then fractionating.

The solids were recrystallized from a mixture of benzene and ethyl acetate.

N-dimethyl acetamide was obtained by heating dimethylamine hydrochloride with excess alkali, drying with pellets of potassium hydroxide and passing into a solution of acetyl chloride in benzene. The precipitated dimethylamine hydrochloride was filtered off and the filtrate fractionated.

In all cases the density values refer to water at 4° C. The values of A and B refer to Sheppard's equation and to the linear part of the curve obtained on plotting  $\log 10^3 \eta$  against T. The boiling-point values of  $\eta$  and D, obtained by extrapolation, are given in parentheses.

*Acetyl chloride*.—B. pt. 55.1° C. at 750.8 mm.  $A = -0.6927$ ;  $B = 375.0$ , over the range studied.

Temp. ....	21.1	32.2	38.9	43.0	48.9	55.1
D .....	1.1057	1.0900	1.0807	1.0745	1.0661	(1.0570)
1000 $\eta$ .....	3.820	3.450	3.234	3.120	2.967	(2.820)

Our interpolated densities at 20° (1.1073) and 25° (1.1003) are close to those of Brühl<sup>(9)</sup> and Walden<sup>(10)</sup> respectively, namely 1.1051 and 1.0982.

*Propionyl chloride*.—B. pt. 79.7° C. at 762.4 mm.  $A = -0.7289$ ;  $B = 412.0$ , over the range studied.

Temp. ....	20.5	36.1	47.3	56.9	64.1	70.9	79.7
D .....	1.0616	1.0396	1.0240	1.0107	1.0004	0.9906	(0.9781)
1000 $\eta$ .....	4.729	4.019	3.614	3.314	3.112	2.946	(2.749)

Our evaluated densities at 20° (1.0622) and 25° (1.0551) compare favourably with those of Brühl and Walden, namely 1.0646 and 1.0508 respectively.

*n-Butyryl chloride*.—B. pt. 100–100.3° C. at 759.8 mm.  $A = -0.7348$ ;  $B = 438.1$ , over the range studied.

Temp. ....	20.6	44.9	59.6	75.6	85.4	95.3	100.3
D .....	1.0263	0.9970	0.9785	0.9575	0.9446	0.9310	(0.9241)
1000 $\eta$ .....	5.735	4.408	3.823	3.322	3.073	2.849	(2.750)

Our calculated density at 20° (1.0270) agrees closely with that given by Brühl, namely 1.0277.

*Succinyl chloride*.—B. pt. 190° C. with decomposition. M. pt. 16.7° (Sugden gives 18.5°)<sup>(11)</sup>.  $A = -0.9089$ ;  $B = 670.6$  above 41°.

Temp. ....	20.3	41.55	52.0	61.65	71.6	82.8
D .....	1.3771	1.3507	1.3375	1.3257	1.3141	1.3002
1000 $\eta$ .....	25.27	16.75	14.23	12.45	10.89	9.45
Temp. ....	100.8	119.2	139.1	159.7	170.0	190.0
D .....	1.2783	1.2552	1.2300	1.2035	1.1900	(1.1638)
1000 $\eta$ .....	7.656	6.313	5.224	4.365	4.018	(3.463)

Above 170° bubbles formed in the viscometer. Our densities agree closely with those of Sugden and Garner.

*Acetamide*.—B. pt. 222.8° at 752.3 mm.  $A = -1.2728$ ;  $B = 921.7$ , above 170°.

Temp. ....	91.1	111.8	131.7	151.2	170.3	190.8	210.0	222.8
D .....	0.9892	0.9711	0.9538	0.9369	0.9202	0.9019	0.8847	(0.8732)
1000 $\eta$ .....	21.82	14.61	10.56	8.039	6.405	5.180	4.320	(3.857)

Our densities agree closely with those of Walker and Johnson<sup>(12)</sup> at 100° and of Dunstan and Mussell<sup>(13)</sup> at 105° and 120°, but our viscosities at the two latter temperatures are higher.

*Propionamide*.—B. pt. 213.1° C. at 759.7 mm.  $A = -1.4187$ ;  $B = 977.8$ , above 170°.

Temp. ...	91.7	112.9	131.3	150.2	170.0	189.8	199.9	209.2	213.1
D .....	0.9438	0.9260	0.9102	0.8936	0.8754	0.8573	0.8481	0.8393	(0.8356)
1000 $\eta$ ....	20.92	13.98	10.22	7.849	6.135	4.947	4.460	4.066	(3.914)

Our viscosities at 105° and 120° are higher than those of Dunstan<sup>(13)</sup>.

*Butyramide*.—B. pt. 216.0° C. at 760.2 mm.  $A = -1.7994$ ;  $B = 1166.6$ , over the range studied.

Temp. ....	121.0	139.4	159.5	176.4	190.1	205.2	216.0
D .....	0.8869	0.8722	0.8555	0.8410	0.8292	0.8164	(0.8072)
1000 $\eta$ .....	14.55	10.72	7.916	6.266	5.242	4.366	(3.858)

*Urethane*.—B. pt. 183.7° C. at 749.7 mm.  $A = -1.5463$ ;  $B = 973.6$  above 80°.

Temp. ...	63.6	86.3	104.9	126.0	145.0	155.2	165.0	175.1	183.7
D .....	1.0462	1.0238	1.0051	0.9834	0.9643	0.9541	0.9440	0.9337	(0.9250)
1000 $\eta$ ...	24.05	14.57	10.69	7.802	6.040	5.311	4.743	4.234	3.853

Our viscosities are again higher at 105° and 120° than those of Dunstan<sup>(13)</sup>, although our densities are almost identical. Stuckgold's viscosities at 70° (18.05) and 80° (14.56) are lower than our interpolated values, 19.6 and 15.7 respectively.

*Dimethyl acetamide*...B. pt. 167.5–168.0° C. at 758.9 mm.  $A = -0.9768$ ;  $B = 670.6$ , above 52°.

Temp. ...	20.4	52.4	72.8	92.2	111.1	132.6	150.7	160.0	168.3
D .....	0.9847	0.9541	0.9339	0.9151	0.8966	0.8757	0.8576	0.8483	0.843
1000 $\eta$ ....	21.41	12.25	9.207	7.244	5.883	4.751	4.032	3.724	3.500

### References.

- (1) Sheppard, J. *Rheology*, i. p. 208 (1930).
- (2) Friend and Hargreaves, *Phil. Mag.* xxxiv. pp. 643, 810 (1943).
- (3) Gibling, J. *Chem. Soc.* p. 299 (1941).
- (4) Sugden, 'The Parachor and Valency,' p. 41 (1930).
- (5) Friend, *Trans. Faraday Soc.* xxxi. p. 542 (1935).
- (6) See discussion in Sidgwick's 'Organic Chemistry of Nitrogen,' p. 143 (1937).
- (7) Friend and Hargreaves, *Phil. Mag.* xxxv. p. 62 (1944).
- (8) Hautsch, *Ber.* lxiv. p. 661 (1931).
- (9) Brühl, *Annalen*, cciii. p. 11 (1880).
- (10) Walden, *Zeitsch. physikal. Chem.* lv. p. 212 (1906).
- (11) Sugden and Garner, *J. Chem. Soc.* p. 2877 (1927).
- (12) Johnson, *J. Chem. Soc.* lxxxvii. p. 1598 (1905).
- (13) Dunstan and Mussell, *J. Chem. Soc.* xcvii. p. 1935 (1910).
- (14) Thorpe and Rodger, *Phil. Trans.* clxxxv. p. 570 (1895).

The Technical College,  
Birmingham.



XXIV. *The Field of a Charged Semi-Infinite Rectangular Conductor Parallel to an Earthed Infinite Plane Conductor or the Flow from a Thick-Walled Jet.*

By N. DAVY, University College, Nottingham\*.

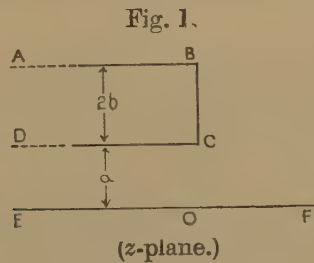
[Received October 30, 1944.]

ABSTRACT.

The problem is a two-dimensional one, which can be regarded as the determination of the electric field of either a charged semi-infinite rectangular conductor parallel to an infinite plane conductor at potential zero, or of two such conductors parallel to each other at equal but opposite potentials. In hydrodynamics it is the problem of the flow from a two-dimensional thick-walled jet into an infinite space. There are corresponding problems in heat flow, magnetism and diffusion. Schwarz-Christoffel transformations involving elliptic functions are obtained, and equations to equipotentials and lines of force are derived. Electric intensities, surface densities, charges and capacities are calculated.

§ 1. *Statement of the Problem and the First Transformation.*

THE first object of this paper is to determine the values of various electrical quantities connected with the two-dimensional system, formed by a semi-



infinite rectangular charged conductor placed alongside and parallel to an infinite plane conductor at potential zero. The system is represented in section in fig. 1, forming the  $z$ -plane. The line ABCDEOF is regarded as the boundary of a polygon, to be transformed into the real axis of the  $t$ -plane. To carry out a Schwarz-Christoffel transformation let A correspond to  $t = -\infty$ , B to  $t = 0$ , C to  $t = \alpha$  where  $\alpha < +1$ , D and E to  $t = +1$ , and F to  $t = +\infty$ . Then the differential transformation is

$$dz = \frac{C_1 t^{\frac{1}{2}} (t - \alpha)^{\frac{1}{2}}}{(t - 1)} dt, \dots \dots \dots (1)$$

---

\* Communicated by Professor L. F. Bates.

which transforms the region external to ABCDEF into the upper half of the  $t$ -plane. CD and OE are parallel straight lines meeting at a common point D or E at infinity. Since CD and OE include an angle zero, their point of intersection contributes the factor  $t-1$  to the denominator of equation (1). To integrate, introduce elliptic functions by writing

$$t = \frac{1}{\text{sn}^2 u} \text{ and } \alpha = k^2, \quad \dots \quad (2)$$

where  $k$  is to be the modulus of the elliptic functions. Then

$$dt = - \frac{2 \text{cn } u \text{ dn } u \text{ du}}{\text{sn}^3 u}, \quad \dots \quad (3)$$

and

$$dz = - \frac{2 C_1 \text{dn}^2 u \text{ du}}{\text{sn}^3 u \text{cn } u} = \frac{2 C_1 (1 - k^2 \text{sn}^2 u) \text{ du}}{\text{sn}^3 u \text{cn } u} \quad \dots \quad (4)$$

$$= - \frac{2 C_1 \cdot \text{du}}{\text{sn}^3 u \text{cn } u} + \frac{2 C_1 k^2 \text{ du}}{\text{sn } u \text{cn } u} \quad \dots \quad (5)$$

TABLE I.

Point.	$x$ .	$y$ .	$t = \frac{1}{\text{sn}^2 u}$ .	$\text{sn } u$ .	$\text{cn } u$ .	$\text{dn } u$ .	$p$ .	$q$ .	$u$ .
A	$-\infty$	$a+2b$	$-\infty$	0	1	1	0	0	0
B	0	$a+2b$	0	$\infty=1$	$-i$	$-ik$	0	$K'$	$iK'$
C	0	$a$	$\alpha=k^2$	$\frac{1}{k}$	$\frac{-ik'}{k}$	0	$K$	$K'$	$K+iK'$
D or E	$-\infty$	$a$	1	1	0	$k'$	$K$	0	$K$
F	$+\infty$	0	$+\infty$	0	1	1	0	0	0

On integrating this we get

$$z = -2C_1 \left[ -k^2 \left\{ \frac{\text{dn } u}{2(1-\text{dn}^2 u)} + \frac{1}{4} \log \frac{\text{dn } u + 1}{\text{dn } u - 1} \right\} + k'^2 \left\{ \log \frac{1-\text{dn } u}{\text{sn } u} - \frac{1}{k'} \log \frac{\text{dn } u - k'}{\text{cn } u} \right\} \right] + C_2 \quad \dots \quad (6)$$

$u$  is in general complex, so we may write  $u=p+iq$  where  $p$  and  $q$  are real. Table I. exhibits a set of values of  $p$ ,  $q$  and  $u$  at various points along the edges of the two conductors. The width of the semi-infinite electrode is  $2b$  cm. and that of the gap OC is  $a$  cm.

To determine the constants  $C_1$  and  $C_2$  of equation (6), values of  $x$ ,  $y$  and  $u$ , for the points B and C in fig. 1, are inserted in (6). At the point B (see Table I.)  $z=i(a+2b)$ ,  $u=iK'$ , giving

$$i(a+2b) = -2C_1 \left\{ k'^2 \log k + \frac{ik'^2\pi}{2} - k' \log k \right\} + C_2 \quad \dots \quad (7)$$

At the point C,  $z=ia$  and  $u=K+iK'$ , giving

$$ia=-2C_1\left\{k'^2\log k-\frac{ik^2\pi}{4}+\frac{ik'\pi}{2}-k'\log k\right\}+C_2. \quad (8)$$

Subtracting (8) from (7) we get, after dividing by  $2i$ ,

$$b=\frac{-C_1\pi}{2}\left\{k'^2-k'+\frac{k^2}{2}\right\}. \quad (9)$$

$$=-\frac{C_1\pi(k'-1)^2}{4}. \quad (10)$$

We also express  $a$  in terms of  $C_1$  if we apply the fact that, in passing from D to E,  $z$  changes by an amount  $-ia$ , and  $t$  passes through the value  $t=+1$ . To use this, integrate equation (1) for values of  $t$  close to  $t=+1$ . The approximate form of (1) is then

$$dz=C_1\frac{(1-\alpha)^{\frac{1}{2}}dt}{t-1}, \quad (11)$$

which is derived from (1) by writing  $t=+1$  in the terms in the numerator but not in the denominator. On integrating (11) we get

$$z=C_1(1-\alpha)^{\frac{1}{2}}\log(t-1)+C_2. \quad (12)$$

Write  $t-1=re^{i\phi}$ .

Then  $z=C_1(1-\alpha)^{\frac{1}{2}}\{\log r+i\phi\}+C_2. \quad (13)$

In passing from D to E the change in  $z$  is  $-ia$ , and as  $t$  passes through the value  $+1$ ,  $r$  remains unaltered and  $\phi$  changes from  $\pi$  to zero. Hence, from (13)

$$-ia=C_1(1-\alpha)^{\frac{1}{2}}\times-i\pi$$

$$\text{and } C_1=\frac{a}{\pi(1-\alpha)^{\frac{1}{2}}}=\frac{a}{\pi(1-k^2)^{\frac{1}{2}}}=\pm\frac{a}{\pi k'}. \quad (14)$$

Taking the negative value of  $C_1$ , which gives positive real values of  $k$ , when (10) and the next equation, (15) are combined, we get

$$a=-\pi k'C_1. \quad (15)$$

Combining (15) and (8) we get

$$C_2=\frac{2a}{\pi}(1-k')\log k+\frac{ik^2a}{2k'}. \quad (16)$$

Dividing (10) by (15) gives

$$\frac{b}{a}=\frac{(k'-1)^2}{4k'}. \quad (17)$$

Hence, when a definite numerical ratio of  $b$  to  $a$  has been selected,  $k'$  and  $k$  are easily found.

$$\text{Let } \frac{b}{a}=n. \quad (18)$$

$$\text{Then } k'^2-(4n+2)k'+1=0. \quad (19)$$

$$k'=(2n+1)\pm\sqrt{4n^2+4n}. \quad (20)$$

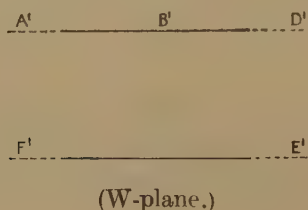
When numerical values of  $k$  and  $k'$  have been found, use can be made of tables of elliptic functions, such as the Smithsonian Mathematical Formulæ and Tables of Elliptic Functions by Adams and Hippisley, 1922.

## § 2. Transformation involving Potentials, Equipotentials and Lines of Force.

To introduce the electrical properties of the system, the  $W$ -plane must now be considered. In this plane the figure consists of two parallel straight lines,  $W=U+iV_0$  and  $W=U$  (see fig. 2). The line  $W=U+iV_0$  corresponds to the edge ABCD in fig. 1 and is therefore called  $A'D'$ , and the line  $W=U$  corresponds to the edge EF in fig. 1, and is called  $E'F'$ . Thus the point  $A'$  corresponds to  $t=\infty$ ,  $D'$  and  $E'$  to  $t=+1$  and  $F'$  to  $t=-1$ . The differential equation, transforming the polygon  $A'D'E'F'$  to the real axis in the  $t$ -plane, is

$$dW = \frac{C_3 dt}{t-1}, \quad \dots \quad (21)$$

Fig. 2.



the factor  $t-1$  in the denominator being contributed by the point  $D'$  or  $E'$ , where the internal angle of the polygon is zero. On integrating we get

$$W = C_3 \log(t-1) + C_4 \quad \dots \quad (22)$$

Let the point  $B'$ , where  $W=iV_0$ , correspond to  $B$ , *i. e.* to  $t=0$ . Then

$$iV_0 = C_3 \log(-1) + C_4,$$

and since  $-1 = e^{i\pi}$ ,  $\log(-1) = i\pi$ , and  $iV_0 = i\pi C_3 + C_4$ . (23)

Since, in passing from  $D$  to  $E$ ,  $W$  changes by  $-iV_0$  and

$$\log(t-1) = \log re^{i\phi} \text{ changes by } -i\pi,$$

$$-iV_0 = -i\pi C_3 \quad C_3 = \frac{V_0}{\pi}, \quad C_4 = 0. \quad \dots \quad (24)$$

Thus the required transformation, obtained by substituting these values in (22), is

$$W = \frac{V_0}{\pi} \log(t-1) \quad \dots \quad (25)$$

which gives

$$t = 1 + e^{\frac{\pi W}{V_0}}; \quad \dots \quad (26)$$



If in (25) we put  $t = \frac{1}{\operatorname{sn}^2 u}$  as in equation (2), (25) becomes

$$W = \frac{V_0}{\pi} \log \left( \frac{1}{\operatorname{sn}^2 u} - 1 \right) = \frac{V_0}{\pi} \log \frac{\operatorname{cn}^2 u}{\operatorname{sn}^2 u} = \frac{2V_0}{\pi} \log \frac{\operatorname{cn} u}{\operatorname{sn} u} \quad (27)$$

Writing  $u = p + iq$ , this gives

$$W = \frac{2V_0}{\pi} \log \frac{\operatorname{cn}(p+iq)}{\operatorname{sn}(p+iq)}, \quad (28)$$

which, on expanding, becomes

$$W = \frac{2V_0}{\pi} \log \frac{\operatorname{cn} p \operatorname{cn} iq - \operatorname{sn} p \operatorname{dn} p \operatorname{sn} iq \operatorname{dn} iq}{\operatorname{sn} p \operatorname{cn} iq \operatorname{dn} iq + \operatorname{cn} p \operatorname{dn} p \operatorname{sn} iq}, \quad (29)$$

all to modulus  $k$ . On transforming functions of  $iq$  to modulus  $k$  into functions of  $q$  to modulus  $k'$ , (29) becomes

$$W = \frac{2V_0}{\pi} \log \frac{\operatorname{cn}(p, k) \operatorname{cn}(q, k') - i \operatorname{sn}(p, k) \operatorname{dn}(p, k) \operatorname{sn}(q, k') \operatorname{dn}(q, k')}{\operatorname{sn}(p, k) \operatorname{dn}(q, k') + i \operatorname{cn}(p, k) \operatorname{dn}(p, k) \operatorname{sn}(q, k') \operatorname{cn}(q, k')}. \quad (30)$$

This can be written in the form

$$W = \frac{2V_0}{\pi} \log \frac{A-iB}{C+iD} = \frac{2V_0}{\pi} \log \frac{\{AC-BD-i(AD+BC)\}}{C^2+D^2}. \quad (31)$$

This can be expressed more simply in the forms

$$W = \frac{2V_0}{\pi} \log (L-iM) = \frac{2V_0}{\pi} \log R e^{i\theta}, \quad (32)$$

where

$$R = (L^2 + M^2)^{\frac{1}{2}} \text{ and } \tan \theta = \frac{M}{L}$$

Further,

Further,

$$W = U + iV = \frac{2V_0}{\pi} \log R - i \frac{2V_0 \theta}{\pi}. \quad (33)$$

Thus equipotentials are given by  $V = \text{const.}$ , i. e.

$$\theta = \text{const.} \quad (34)$$

and lines of force by  $U = \text{const.}$ , i. e.

$$R = \text{const.} \quad (35)$$

If  $\theta$  is const.,

$$\tan \theta = \text{const.} \quad (36)$$

that is,

$$\frac{AD+BC}{AC-BD} = \text{const.} \quad (37)$$

Let

$$\left. \begin{aligned} \operatorname{sn}(p, k) &= s_1, & \operatorname{sn}(q, k') &= s_2. \\ \operatorname{cn}(p, k) &= c_1, & \operatorname{cn}(q, k') &= c_2. \\ \operatorname{dn}(p, k) &= d_1, & \operatorname{dn}(q, k') &= d_2. \end{aligned} \right\} \quad (38)$$

A, B, C, and D can be expressed in terms of these quantities by means of equations (31) and (30), and then, using the values thus obtained, along with (38), in (37) the latter becomes, after some rearrangement,

$$c_1^2 d_1 c_2^2 s_2 + s_1^2 d_1 s_2 d_2^2 = \text{const.} (s_1 c_1 c_2 d_2 - s_1 c_1 d_1^2 s_2^2 c_2 d_2) . \quad (39)$$

This is the equation to the whole family of equipotentials. Before an equipotential curve can be drawn in fig. 1, we must be able to find pairs of values of  $p$  and  $q$  which satisfy (39) for a given value of the constant. Suppose that numerical values of  $k$  and  $k'$  have been selected by the aid of (17) and (20). Then a value of  $p$  is chosen at will, and numerical values of  $s_1$ ,  $c_1$  and  $d_1$  for  $p$  to modulus  $k$  are taken from the tables. Inserting these in (39), it reduces to

$$M c_2^2 s_2 + N s_2 d_2^2 = \text{const.} c_2 d_2 (O - P s_2^2), \quad . \quad . \quad . \quad (40)$$

where M, N, O and P are constants. Expressing  $c_2$  and  $d_2$  in terms of  $s_2$ , and squaring to eliminate roots, (40) becomes an equation of the eighth degree in  $s_2$ , which is not a biquadratic in  $s_2^2$ . Hence it does not yet seem possible to get  $s_2$  and hence  $q$ , and the value of  $u$  which by insertion in (6) would give a point on an equipotential. On the other hand, if we treat equation (35), the equation to a line of force, in the same manner as (37), it reduces to

$$c_1^2 c_2^2 + s_1^2 d_1^2 s_2^2 d_2^2 = \text{const.} (s_1^2 d_2^2 + c_1^2 d_1^2 s_2^2 c_2^2). \quad . \quad . \quad (41)$$

This reduces to a soluble biquadratic in  $s_2$  when we insert chosen numerical values of  $s_1$ ,  $c_1$  and  $d_1$ , and the constant in it, so that  $q$  and  $u$  can be obtained and co-ordinates of points on lines of force can be obtained by inserting values of  $u$  in (6).

### §3. *Electric Intensities, Surface Densities and Total Charges.*

The electric intensity at any point is

$$F = \left| \frac{dW}{dz} \right| . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Hence, to find F, differentiate equation (27). This gives

$$\frac{dW}{du} = - \frac{2V_0}{\pi} \cdot \frac{dn \, u}{sn \, u \, cn \, u} . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Also from (4)

$$\frac{dz}{du} = - 2C_1 \frac{dn^2 u}{sn^3 u \, cn \, u} . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Dividing (43) by (44), we get

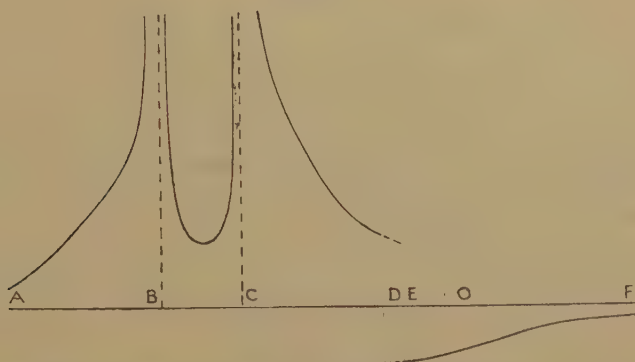
$$\left| \frac{dW}{dz} \right| = F = \frac{V_0}{\pi C_1} \frac{sn^2 u}{dn \, n} = - \frac{V_0 k'}{a} \left| \frac{sn^2 u}{dn \, u} \right| . \quad . \quad . \quad . \quad . \quad (45)$$

$\sigma$ , the surface density of charge at any point

$$= \frac{F}{4\pi} = - \frac{V_0 k'}{4\pi a} \left| \frac{sn^2 u}{dn \, u} \right| . \quad . \quad . \quad . \quad . \quad (46)$$

As a check on this result it may be noted that, as expected,  $\sigma$  becomes infinite at the corner B in fig. 1, where by Table I.  $\text{sn } u = \text{an infinite quantity}$  I, and  $\text{dn } u = -ikI$ , so that  $\left| \frac{\text{sn}^2 u}{\text{dn } u} \right|$  is infinite. Also, at the corner C,  $\sigma$  is also infinite, since  $\text{sn } u = 1/k$  and  $\text{dn } u = 0$ . Equation (46) also predicts that  $\sigma$  is zero at the point A, which is an infinite distance away along BA, since there  $u=0$ ,  $\text{sn } u=0$ ,  $\text{dn } u=1$ , and the same at F. At D or E, where  $t=+1$ ,  $u=K$ ,  $\text{sn } u=1$ ,  $\text{dn } u=k'$  and  $\sigma = \frac{V_0}{4\pi a k'}$ . In fig. 3, graphs representing the variation of  $\sigma$  along the two conductors are drawn.

Fig. 3.



(Variation of surface density  $\sigma$  along the edges of the conductors.)

$\sigma$  is proportional to the height above the horizontal axis at any point. It is negative along EF.

The total charge per cm. depth residing on any length of a conductor is

$$Q = \int_{z_1}^{z_2} \sigma \cdot dz = \frac{1}{4\pi} \int_{z_1}^{z_2} \left| \frac{dW}{dz} \right| dz = \frac{1}{4\pi} (W_2 - W_1). \quad (47)$$

$W$  is given by equation (27). Hence

$$\begin{aligned} Q &= \frac{1}{4\pi} \cdot \frac{2V_0}{\pi} \cdot \left[ \log \frac{\text{cn } u}{\text{sn } u} \right]_1^2 \\ &= \frac{V_0}{2\pi^2} \log \frac{\text{sn } u_1 \text{cn } u_2}{\text{sn } u_2 \text{cn } u_1}, \quad \dots \dots \dots (48) \end{aligned}$$

where  $u_1$  and  $u_2$  are the values of  $u$  at the ends of the portion of conductor considered. On substituting in (48) the appropriate values of  $u$  for various ranges, the following results are obtained: Range A to B in fig. 1,  $Q$  is infinite; range B to C,

$$Q = \frac{V_0}{2\pi^2} \log k' \text{ in e.s.u. ;}$$

range C to D,  $Q$  is infinite; range E to F,  $Q$  is infinite.





in the direction of increasing  $z$ . Here  $\chi_2$  and  $\chi_1$  are the mass susceptibilities of the body and surrounding medium, respectively.

By (45) 
$$F = -\frac{V_0 k'}{\alpha} \left| \frac{\operatorname{sn}^2 u}{\operatorname{dn} u} \right|.$$

Hence 
$$\frac{dF^2}{dz} = \frac{dF^2}{du} \cdot \frac{du}{dz} = \frac{\pi V_0^2 k'^3 \operatorname{sn}^6 u \operatorname{cn}^2 u (1 + \operatorname{dn}^2 u)}{a^3 \operatorname{dn}^5 u} \dots \quad (52)$$

by equations (45) and (4). This expression and the mechanical force on the small body are therefore infinite when the body is placed near the corner C where  $\operatorname{dn} u = 0$ , and when it is placed near the corner B where  $\operatorname{sn} u = \infty = I$ ,  $\operatorname{cn} u = -iI$ ,  $\operatorname{dn} u = -ikI$ . The force is zero when the body is placed near A or F where  $\operatorname{sn} u = 0$ , and near D or E where  $\operatorname{cn} u = 0$ .

### 7. *Two Parallel Electrodes or a Thick-walled Jet.*

The field considered above will not be changed at any point when we replace the infinite wall by the electrical image of the semi-infinite conductor in the wall. There is symmetry about the plane of the wall. The image is a second semi-infinite conductor of equal but opposite potential to the first. Hence the foregoing theory solves the problem of finding the electrical quantities connected with a pair of equal semi-infinite parallel rectangular conductors, with equal but opposite potentials. The results expressed in equations (6), (25), (27), (45), (47) and (48) still hold, but the capacity between the end BC and its image is half that given by (49), *i. e.* it is  $\frac{1}{4\pi^2} \log k'$ , and the resistance between

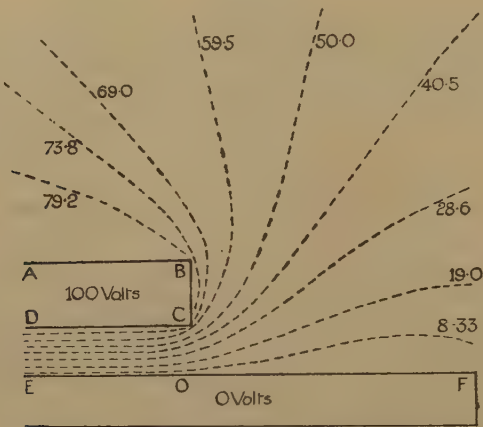
BC and its image is doubled, *i. e.* it is  $\frac{\rho\pi}{\log k'}$ . In hydrodynamics this case is that of the flow from a thick-walled jet into an infinite space. The curves  $V = \text{const.}$  of equations (34), (36), (37) and (39) are the stream lines, and the curves  $U = \text{const.}$  {equations (35) and (41)} are the lines of equal velocity potential. The modulus of  $-\frac{dz}{dW}$  is the reciprocal of the resultant velocity at any point.

### 8. *A Supporting Experiment.*

An electrical experiment was planned to obtain the equipotential curves of the system of fig. 1. The theoretical electrodes are infinitely long so, of course, they are represented by finite ones. The infinite wall EF was represented by a strip of brass, cut off at F. ABCD was another rectangular strip of brass. These brass electrodes were placed in position as in fig. 4, resting on a uniform sheet of blotting-paper soaked in copper sulphate solution. By means of a beat frequency oscillator, an alternating potential difference was applied between the electrodes. By means of pointed metal probes connected to a voltmeter or headphones, points on equipotential curves were located, and their root mean

square potentials relative to the electrode EF measured approximately. The points on the curves were registered by means of carbon paper pressing on white paper below the blotting-paper. The family of equipotentials, with their R.M.S. potentials represented as percentages of the potential difference between the two electrodes, is shown in fig. 4.

Fig 4.



Experimentally obtained equipotentials.

It will be noticed that as a result of the finite length of EF the equipotential marked 8.33 shows a distinct curl towards EF on the right. Such an experimental method as this may be recommended as a quick way of getting equipotentials in these cases. The labour in calculating the co-ordinates of points from equations such as (39) is always laborious and often impossible.

ERRATA.

Dr. N. W. McLachlan’s paper on “Heat Conduction in Elliptical Cylinder, etc.” *Phil. Mag.* ser. 7, vol. xxxvi, pp. 600–609 (Sept. 1945).

The author asks that the following corrections be made :—

Page 603 :

For  $\int_0^{2\pi} ce_{2n}^2(\eta, k_{2n,m}^2) \cos 2\eta d\eta = 0$  read  $\int_0^{2\pi} ce_{2n}^2(\eta, k_{2n,m}^2) \cos 2\eta = \pi \Theta_{2n},$

where  $\Theta_{2n} = 2A_0A_2 + \sum_{r=1}^{\infty} A_{2r}A_{2r+2}.$

For  $\cosh 2\xi$  read  $[\cosh 2\xi - \Theta_{2n}]$  in the following places :—r.h.s. of (6) § 4, and in the denominators of (7), (8), § 4, (2) § 8, (3) § 9. When the ellipse degenerates to a circle,  $\Theta_{2n} \rightarrow 0.$

Appendix I :—In (3) for  $\zeta$  read  $\zeta_{n,m}$  ; in (4) for  $k_{n,m}^2$  read  $k_{p,r}^2$ , and for the r.h.s. of (8) read  $\neq 0.$

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXV. *Crystal Habit and Internal Structure.*—II.

By A. F. WELLS, M.A., Ph.D.\*

[Received October 1, 1945.]

*Introduction.*

IN Part I. we mentioned some of the factors which influence the habits of crystals, and the equilibrium shape of crystals was discussed. In the present paper we shall examine some aspects of the relation between the face development and the atomic structure of a crystal. During the past few years a considerable number of papers on this subject has been published, and it is important that this work should be viewed in proper perspective in relation to the more general problem which was outlined in Part I.

*The Relation between Bravais Lattice, Space-group and Face Development.*

In a crystal the structural units are arranged in conformity with one of the fourteen Bravais lattices. Bravais suggested that the face development of a crystal was closely related to its fundamental lattice, the faces being parallel to the net planes in the lattice and those planes most closely packed with lattice points being the most important forms† on the crystal. Friedel (1904, -05, -07) showed that this is often true. The Bravais "law" implies that the Miller indices of crystal faces will be small integers, and it is well known that forms with high indices are rare. It was not to be expected that the law would be universally valid for it took account only of the internal structure of the crystal, and it was realized that external factors such as the rate of crystallization and the presence of impurities in the solution could affect the habits of crystals.

The Bravais law was elaborated by Donnay and Harker (1937), who showed that not only the type of lattice but also the presence of screw axes and glide planes (other than those inherent in the lattice) should be taken into account in correlating face development with internal structure. The Bravais law states that the importance of a form  $\{hkl\}$  is proportional to the reticular density (number of lattice points in unit area) or inversely proportional to the reticular area  $S_{hkl}$ . Since  $S_{hkl} \cdot d_{hkl} = V$ , where  $d_{hkl}$  is the spacing of successive planes  $hkl$  and  $V$

---

\* Communicated by the Author.

† A form is a set of crystallographically equivalent faces related to one another by the symmetry elements.

is the volume of the unit cell, the importance of a form should be directly proportional to the interplanar spacing  $d_{hkl}$ . Donnay pointed out that since screw axes and glide planes involve translations, their presence alters the effective spacings of planes perpendicular to the former and of certain of the planes perpendicular to the latter. For example, a  $3_1$  axis changes the effective spacing of the (0001) plane from  $c$  to  $c/3$ , and since this is the only plane affected, the importance of the form {0001} relative to others is altered. It is evident that the planes affected in this way are just those of which certain orders of reflexion of X-rays are absent owing to destructive interference (the systematic absences given in space-group tables).

If the importance of a form  $\{hkl\}$  is proportional to the spacing  $d_{hkl}$  it is possible to draw up for any crystal a list of planes arranged in order of decreasing  $d_{hkl}$ , and the position of a given plane in this "morphological aspect" (M.A.) should express the importance of the form relative to that of others. The Donnay-Harker extension of the Bravais law requires that instead of using in all cases the simplest Miller indices we use the multiple indices of the lowest order of X-ray reflexion compatible with the space-group symmetry. In propounding the new law the authors claimed only that it was an "improved approximation," for it is, of course, subject to the same limitations as regards external factors as was the original law of Bravais. In addition, Donnay and Harker pointed out that if the space-group symmetry were the only influential factor then all cubic crystals, for example, with the same space-group would exhibit the same face development, whereas this is not so.

A morphological aspect is defined as "a possible list of decreasingly important forms for any set of axial elements (axial ratios and inter-axial angles) in any given crystal system." Corresponding to the 230 space-groups there are 97 such morphological aspects (Donnay and Harker, 1940). Only in the cubic system can the planes be listed for a given M.A. without reference to a particular crystal; in other systems the axial elements affect the order of the planes. In order to determine the space-group from the face development, the M.A. drawn up from observations on the crystals is compared with the above theoretical M.A.'s. In the case of minerals the experimental M.A. is drawn up in a standard way (Tremblay, 1942) by making a statistical survey of the frequency of occurrence and relative sizes of various forms on a large number of crystals from different localities. The presence of a plane as a cleavage was originally taken into account in assessing its position in the M.A., but this practice was discontinued in later work. Donnay and others have shown that in a number of cases there is good agreement between the observed face development and that predicted from the space-group, and that in some cases the space-group could be deduced (or confirmed) from morphological studies (Donnay, 1939). Crystals studied in this way include rhombic sulphur, pyrites, spinel and analcime (Donnay and Harker, 1937), stephanite (Taylor, 1940), garnet (Donnay



and Faessler, 1941), scheelite (Donnay, 1942), goethite and lepidocrocite (Peacock, 1942), carborundum (Donnay, 1943), and muscovite (Peacock and Ferguson, 1943).

Now because relationships between space-groups and face developments such as that proposed by Donnay are admittedly of restricted validity, a critical examination of the "exceptions" should indicate some of the other factors determining the stability of crystal faces. We have to enquire (1) how we should expect reticular densities, as defined above, to be related to the stabilities of crystal faces, (2) whether the Donnay-Harker treatment gives a true estimate of relative reticular densities, and (3) what is the significance of the M.A. derived from a survey of observed face developments?

### *The Significance of Reticular Density.*

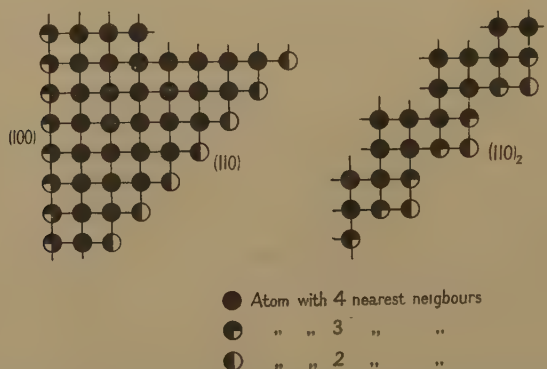
This point can only be discussed in detail with reference to particular crystal structures. Reticular density has been defined as the number of lattice points in unit area of a particular plane. We shall show, however, that the criterion of absolute coplanarity is not applicable when estimating the relative stabilities of different faces of a crystal (see the structure of diamond, below), and that there is no simple way of deciding which atoms shall be counted as "surface atoms."

It should perhaps be emphasized that a plane section through a crystal structure will seldom give a true picture of the contour of an actual crystal face. Apart from the mosaic structure and the defects of various kinds which are known to exist in the interiors of practically all crystals, we have to recognize that additional surface irregularities will generally occur. Random imperfections which are quite large on the atomic scale may be present on optically plane surfaces—see, for example, the work of Tolanski (1945) on the topography of crystal faces. In this connection reference may be made to a concept introduced by Stranski (1932). Fig. 1 represents a section through a simple cubic lattice (perpendicular to the  $c$ -axis) showing the traces of (100) and (110) planes. When counting the nearest neighbours of any atom we shall neglect two which lie respectively above and below each atom in this section. If we draw the (110) plane in the ordinary way we see that the outermost atoms have only two neighbours, while those in the next layer have four. There is also, however, an infinite set of planes  $(110)_n$  with steps of different sizes. That with double steps,  $(110)_2$ , is shown in fig. 1, and mixed types with steps of different sizes can also be visualized. It will be seen that  $(110)_2$  is intermediate between a (110) and a (100) plane, and this hybrid character becomes more marked as the size of the steps increases. In what follows we shall assume that a crystal face has the simplest possible type of contour, but it must be remembered that this is an idealized picture of an actual crystal face.

By taking account of glide planes and screw axes, Donnay and Harker imply that the "reticular densities" relevant to a discussion of the relative stabilities of crystal faces refer to the numbers of coplanar atoms

per unit area. Their treatment allows only for symmetry operations which move atoms by amounts which are certain sub-multiples of the appropriate cell dimensions, viz.  $1/n$  where  $n$  is 2, 3, 4, or 6, and which therefore lead to extinction of certain X-ray reflexions. However, any movement of an atom out of a plane affects the density of atoms in the plane if this is taken to mean the number of atoms exactly in the plane, though it does not, except in the cases mentioned above, lead to a space-group extinction. It does, of course, affect the intensity of X-rays from the plane, and if the intermediate atoms are suitably placed may lead to complete extinction of certain orders of reflexion. Since such extinctions are dependent on the value of a variable parameter they are not "space-group extinctions" and are not allowed for in the M.A. of Donnay and Harker. In other words, to obtain a true M.A. we should have to consider which atoms have variable parameters, *i. e.*

Fig. 1.



Contours of (100), (110) and (110)<sub>2</sub> faces (see text), showing environments of surface atoms.

whether they occupy special or general positions in the space-group. The fundamental difference between lattice halvings and those due to screw axes and glide planes arises from the fact that a lattice centring affects every atom in the unit cell in the same way, whereas this is not so in the case of symmetry elements additional to those inherent in the lattice. In the latter case the number of equivalent atoms related by the operation of the symmetry element is determined by the position of an atom relative to the symmetry element, which also determines the number of variable parameters required to define their positions. The densities of atoms in certain planes may be greatly enhanced if the parameters of the atoms have certain values, and in such a case the M.A. of Donnay and Harker may bear very little relation to the relative importance of the various forms. This point is very evident in the morphology of crystals containing planar molecules, and may be illustrated by reference to the structures of naphthalene, anthracene and benzoquinone.

*The Morphology of Naphthalene, Anthracene and Benzoquinone.*

These compounds all crystallize in the monoclinic system, space-group  $P2_1/a$ , and their structures and face developments are very similar. The most prominent forms are in all cases  $\{001\}$ ,  $\{110\}$  and  $\{20\bar{1}\}$ . In Table I. the first eight planes of anthracene (Robertson, 1933) and benzoquinone (Robertson, 1935) are listed in order of increasing  $\sin \theta$  (or decreasing  $d_{hkl}$ ), the observed forms being marked with asterisks. These lists form the M.A.'s of Donnay and Harker. The agreement with the observed face development is obviously poor, particularly in the case of benzoquinone. The importance of  $\{20\bar{1}\}$  and the good  $(20\bar{1})$  cleavage are due to the fact that the molecules lie in this plane. Accordingly  $(20\bar{1})$  is the strongest X-ray reflection and the higher orders decrease uniformly in intensity. These two criteria together furnish the best test for coplanarity of atoms.

TABLE I.

Anthracene		Benzoquinone		Anthracene	
$hkl$	$\sin \theta$	$hkl$	$\sin \theta$	$hkl$	$S/\sin \theta$
001*	0.084	001*	0.135	110*	533
011	0.152	110*	0.159	001*	519
11 $\bar{1}$	0.158	011	0.176	20 $\bar{1}$ *	410
110*	0.167	11 $\bar{1}$	0.197		
002	0.168	111	0.222	210	352
		200	0.223		
20 $\bar{1}$ *	0.184	020	0.226		
20 $\bar{2}$ }		20 $\bar{1}$ *	0.238		
11 $\bar{2}$				etc.	

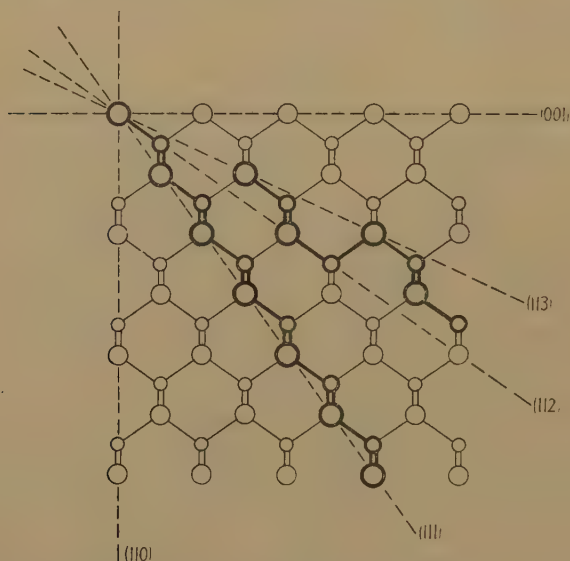
If, therefore, we wished to draw up a true M.A. from the X-ray data it would not be sufficient merely to allow for the space-group absences, as do Donnay and Harker, but also to consider the intensities of the X-ray reflections, thereby taking into account the actual parameters of atoms in general positions. We could, for example, list the *first-order* reflections (others are irrelevant) in order of decreasing  $S/\sin \theta$ , where  $S$  is the geometrical structure factor for the plane (see Table I.). For a first-order reflection  $S$  is a measure of the coplanarity of the atoms, but it does not take account of the distribution of the atoms in the plane. For this reason  $\sin \theta$  is introduced. Unfortunately, however,  $S$  is not a measure of the departures from coplanarity in *absolute* measure but only in terms of the spacing of each particular plane. We shall not pursue this point further as it will soon be evident that we cannot expect to find general relations between space-groups and face developments which will be applicable to all types of crystals.

We mentioned above that the criterion of absolute coplanarity is not of general applicability in estimating the stabilities of crystal faces. This point can be conveniently discussed with reference to the structures of diamond and zinc blende.

*The Morphology of Diamond and Zinc Blende.*

In the unit cell of diamond atoms occupy the two sets of positions:  $C_1$  ( $000$ ,  $0\frac{1}{2}\frac{1}{2}$ ,  $\frac{1}{2}0\frac{1}{2}$ ,  $\frac{1}{2}\frac{1}{2}0$ ) and  $C_2$  ( $\frac{1}{4}\frac{1}{4}\frac{1}{4}$ ,  $\frac{1}{4}\frac{3}{4}\frac{3}{4}$ ,  $\frac{3}{4}\frac{1}{4}\frac{3}{4}$ ,  $\frac{3}{4}\frac{3}{4}\frac{1}{4}$ ), *i. e.*, they form two interpenetrating face-centred cubic lattices related by a translation of ( $\frac{1}{4}\frac{1}{4}\frac{1}{4}$ ). The space-group is  $Fd\bar{3}m$ . In addition to those for a face-centred lattice ( $h, k$  and  $l$  all odd or all even) there are additional halvings due to

Fig. 2.



Projection of the crystal structure of ZnS (or diamond) on  $(\bar{1}10)$ , showing the contours of  $(111)$ ,  $(112)$  and  $(113)$  faces. Zn and S (or  $C_1$  and  $C_2$ ) atoms are distinguished as large and small circles respectively.

the diagonal glide planes ( $h0l$ ,  $h+l=4n$ , etc.). Allowing for these various halvings the M.A. is

$$\begin{array}{cccccc} 111 & 110 & 113 & 001 & 133 & 112, \text{ etc.}, \\ & (220) & & (004) & & (224) \end{array}$$

giving first place to  $(111)$ , which is the dominant form and also a perfect cleavage. The agreement would appear to be satisfactory, but we shall see that  $(111)$  is not the plane most densely packed with *coplanar* atoms, although it comes first in the above series. In the zinc blende structure the Zn and S atoms occupy the above sets of equivalent positions, the space-group is  $F\bar{4}3m$ , and  $(111)$  occupies first place in the M.A. as in the case of diamond. However, the common forms on blende are  $\{110\}$ ,



{111}, and {113}, probably {110} most important, and the cleavage is (110), perfect. These differences between two crystals with the same atomic arrangement call for explanation.

From fig. 2 it will be seen that the S atoms fall exactly *in* some but *between* other planes of Zn atoms. The second column in Table II. gives the numbers of Zn atoms per unit area in the various planes, the third column the type(s) of atom lying *in* the plane, and the fourth column the total numbers of atoms per unit area. It is seen that (110) is the plane most densely populated with atoms (Zn+S), and the next plane in which Zn and S are coplanar is (112). (For these planes the figures in column 4 are twice those in column 2.) The importance of (111) both as a face and a cleavage of diamond shows that the criterion of absolute coplanarity is not the one determining stability (see fig. 2). In spite of the fact that  $C_1$  and  $C_2$  are strictly coplanar for (112) but not for (113), the latter is the more stable face because the contour of a (113) face will obviously be that shown in fig. 2, containing both  $C_1$  and  $C_2$ . The difference between the stabilities of (110) and (111) for diamond and ZnS must be associated with the partial ionic character of the bonds in the latter.

TABLE II.

$hkl$	Zn/ $a^2$	Atoms in Plane	Total Atoms/ $a^2$
111	2.31	2.31 Zn or 2.31 S	2.31
100	2.00	2.00 Zn or 2.00 S	2.00
110	1.41	1.41 Zn + 1.41 S	2.82
113	1.21	1.21 Zn or 1.21 S	1.21
133	0.92	0.92 Zn or 0.92 S	0.92
210	0.89	0.89 Zn or 0.89 S	0.89
112	0.82	0.82 Zn + 0.82 S	1.64

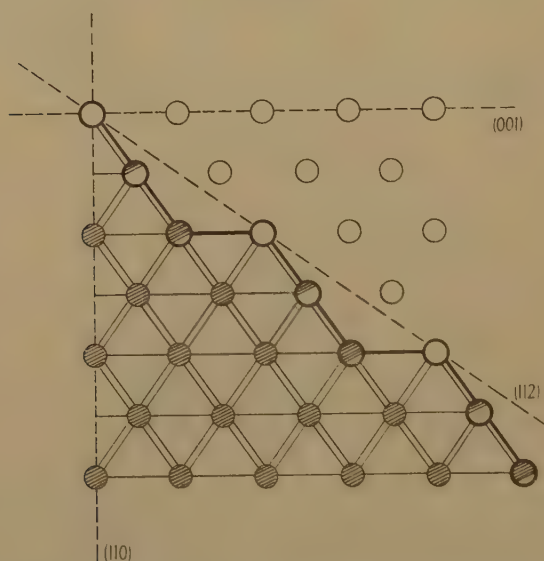
If we count as surface atoms for (111) all those shown (for this plane) as heavy circles in fig. 2, then one-half are forming four and the other half three bonds (a total of  $4.62 \text{ atoms}/a^2$ ), compared with  $2.82/a^2$  all forming three bonds in (110),  $a$  being the side of the unit cell. Thus for diamond (111) is more stable than (110). In the case of ZnS, however, the outer atoms of (111) are all Zn (or all S) forming only three bonds, so that there is a resultant positive or negative charge (compare (111) of NaCl). On the other hand (110) contains equal numbers of Zn and S atoms all with the same residual charge, so that the surface is electrically neutral. A (111) cleavage of ZnS would mean the separation of oppositely charged surfaces (contrast the effect of a (110) cleavage), whereas any cleavage of diamond leaves two neutral carbon surfaces and a (111) cleavage breaks the smallest number of C-C bonds per unit area.

We shall now consider in turn three structures which all have the same space-group,  $Fm3m$ , viz., the face-centred cubic, the rock-salt and the fluorite structures.

*The Face-centred Cubic Structure.*

In this structure (cubic close-packing) every atom has twelve equidistant nearest neighbours. An atom in the surface of the crystal has a smaller number of neighbours, this number depending on the indices of the face. Not only are the atoms in the outermost layer affected but in general also those in layers below the surface. This may be seen from fig. 3, which shows the plan of the structure on  $(\bar{1}10)$  and the traces of certain planes which are perpendicular to the plane of the paper. The outermost atoms in a  $(112)$  face, for example, have only seven nearest neighbours (two are not shown, there being one above and one below the plane of the paper), those in the next layer have nine, and in the next ten.

Fig. 3.



The face-centred cubic structure projected on  $(\bar{1}10)$ , showing the environments of atoms forming a  $(112)$  face.

Only when we come to the fourth layer of atoms below the surface is the environment the same as in the interior of the crystal. These differences are shown by the shading of the circles in fig. 3.

In Table III. the planes in this structure are listed in order of decreasing  $d_{hkl}$ , and the environments of the surface atoms are given. The indices used in calculating the spacings of the planes are those consistent with the conditions for reflection of X-rays from a face-centred lattice, viz.,  $h$ ,  $k$  and  $l$  all odd or all even. The planes are arranged in order of decreasing density of packing of atoms in the planes, *i. e.*, they form the M.A. for the space-group  $Fm\bar{3}m$ .

*The Rock-salt Structure.*

In this structure the co-ordinates of each type of ion are those of cubic close-packing, and in the interior of the crystal each ion is surrounded octahedrally by six of the other kind. Table IV. shows the planes arranged in order of decreasing density of packing of Na or Cl ions. It shows, for example, that (111) is the plane most closely packed with ions of *one kind*. This is not, however, the criterion for stability of a face on a NaCl crystal. It is likely that electrical neutrality is the criterion in the case of an essentially ionic crystal, or alternatively, that the surface ions have

TABLE III.

$hkl$	$d_{hkl}=1/(\bar{h}^2+\bar{k}^2+\bar{l}^2)^{\frac{1}{2}}$	Neighbours of Surface Atom
111	0.577	9
200 (100)	0.500	8
220 (110)	0.353	7 (11)
113	0.301	7 (10)
133	0.229	7 (9) (11)
420 (210)	0.224	6 (9) (11)
224 (112)	0.204	7 (9) (10)

TABLE IV.

$hkl$	Nature of Plane	Total No. of Ions in the Area $a^2$	Number of oppositely charged Neighbours of an Ion in the Face
111	Na or Cl	2.31	3
200	Na+Cl	4.00	5
220	Na+Cl	2.82	4
113	Na or Cl	1.21	3 (5) (5)
133	Na or Cl	0.92	3 (4) (4)
420	Na+Cl	1.79	4 (5)
224	Na+Cl	1.63	3 (5)

the maximum number of neighbours of opposite charge. Table IV. shows the environments of the ions in different faces of a crystal of NaCl, the planes being arranged in the same order as in Table III. The third column gives the total numbers of ions in the area  $a^2$ ,  $a$  being the length (in A. units) of the side of the unit cell. The planes fall into two groups, and the figures in which we are interested are those to the right in column 4. We have, therefore, two possibilities for a compound AX with the rock-salt structure. If the condition for stability of a face is that it contains equal numbers of A and X atoms, then (100) is the most stable face. If, for any reason, the most stable face is one consisting of atoms of one kind

only, then (111) will occur. This seems to apply to the less ionic  $\text{CdO}$ . It appears that this crystal is most stable if its outer surface consists of oxygen atoms, for {111} is the dominant form. The same development will be observed if interaction with the solvent or impurity in the solution stabilises the face consisting of all A or all X ions—see the discussion [below] of the habit of  $\text{NaF}$  from water and  $\text{NaCl}$  from solutions containing  $\text{OH}^-$  ions or  $\text{CO}(\text{NH}_2)_2$ .

### *The Fluorite Structure.*

The natures of the faces of a crystal of  $\text{CaF}_2$  are shown in Table V. Only three planes are given as they are sufficient to illustrate the point we wish to emphasize. It is assumed that the contour of a (111) face would be that indicated in fig. 4, *i.e.*, it is electrically neutral with an outer surface consisting of practically coplanar Ca and F ions. This should be the most stable face if the criterion is, as in the case of  $\text{NaCl}$ , electrical neutrality. If the condition for stability were the greatest density of F ions then (100) should develop. The mineral fluor spar is almost invariably in the form of cubes, and the same development is

TABLE V.

hkl	Nature of Plane	Numbers of Ions in Area $a^2$	Environment of Ion in Surface
111	$\text{Ca} + \text{F}$	$2.3 \text{ Ca} + 2.3 \text{ F}$	$\text{Ca (7F) and F (3Ca)}$
100	$\text{Ca or F}$	$2 \text{ Ca or } 4 \text{ F}$	$\text{Ca (4F) or F (2Ca)}$
110	$\text{Ca} + 2\text{F}$	$1.4 \text{ Ca} + 2.8 \text{ F}$	$\text{Ca (6F) and F (3Ca)}$

shown by crystals grown by the slow diffusion of solutions of  $\text{CaCl}_2$  and  $\text{NH}_4\text{F}$ . The development of (100), instead of the more densely populated and electrically neutral (111) or (110), is in striking contrast to the development of (100) on  $\text{NaCl}$ , and is referred to again later. The perfect (111) cleavage is consistent with the assumption that the cleavages on simple ionic crystals result in the separation of electrically neutral surfaces by the rupture of the smallest possible number of bonds per unit area, or alternatively the separation of the most densely populated electrically neutral planes. The (111) cleavage of  $\text{CaF}_2$  breaks  $4.62$  bonds  $a^2$ , but a (110) cleavage  $5.67$  bonds  $a^2$ . Similarly, a (100) cleavage of  $\text{NaCl}$  breaks  $4$  bonds  $a^2$ , but a (110) cleavage  $5.6$  bonds  $a^2$ . (An equation for the cleavability of ionic minerals proposed by Shappell (1936) fails to account for the cleavages of these two simple crystals, though it has some success with more complex structures.)

In general we should expect some relation between the ease of cleavage parallel to a particular plane and the stability of that plane as a crystal face for crystals throughout which all the bonds are of the same kind ( $\text{NaCl}$ ,  $\text{ZnS}$ , diamond, etc.). Where there are pronounced differences between the types and/or strengths of bonds in different directions in the



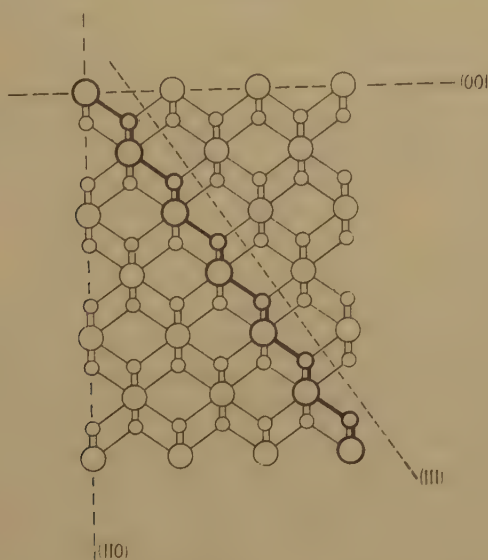
crystal, good cleavages are often observed, as in the case of many crystals with layer structures. It does not follow, however, that these cleavage planes are stable in contact with the saturated solution of the substance in a particular solvent. We have already noted in Part I. that the (001) cleavage plane of pentaerythritol is highly unstable in a saturated aqueous solution.

*Conclusions: Relation between Space-group and Morphology.*

These considerations lead to the following conclusions:—

(1) The use of “reticular densities” is not satisfactory in this connection. The Donnay-Harker treatment implies that only atoms

Fig. 4.



Projection of the structure of  $\text{CaF}_2$  on  $(\bar{1}10)$  showing the contour of a (111) face.  
Large circles represent Ca ions.

which are exactly coplanar are to be considered as lying in a particular plane and contributing to its stability. This criterion of absolute coplanarity is inadmissible, as shown by the morphology of diamond. There would not appear to be any simple way of estimating, directly from the space-group, which atoms should be counted as “surface atoms” for a particular face.

(2) The criterion for stability of a face is not necessarily the density of atoms per unit area regardless of their nature, so that if the crystal contains atoms of more than one kind it is necessary to know which type of atom or combination of atoms forms the most stable surface. This will depend on the nature of the crystal. For an ionic crystal the condition appears to be electrical neutrality, provided that there is no

preferential interaction between the solvent and any of the constituent ionic species (see the discussion of NaF, below). If, on the other hand, there is strong interaction between certain atoms or certain parts of the molecules of a solute with molecules of the solvent, then the most stable faces will tend to be those containing the greatest concentrations of these active atoms or groups, and these are not necessarily the most densely populated planes.

(3) Even within the above limitations, a "morphological aspect" which takes account only of space-group halvings (due to lattice-centring, glide planes and screw axes) does not necessarily give a true indication of the relative importances of faces of different types, though it may give a good approximation in some cases.

(a) The M.A. does not allow for the difference in stability of pairs of parallel faces in cases of hemihedral symmetry, since their indices have the same numerical values (see CuI, below).

(b) If all the atoms in the unit cell are in positions exhibiting the symmetry of the space-group, the M.A. will give a measure of the relative stabilities of the faces *with the proviso* that much higher concentrations of atoms may occur in certain planes owing to the values of the parameters. This is not allowed for if we take account only of the space-group absences, but will show itself in the relative intensities of the X-ray reflexions.

(c) If some of the atoms are in special positions of a space-group other effects may arise. For example, the atoms of one kind may be arranged in accordance with the symmetry of another space-group (often of higher symmetry), as may be seen from the lists of Gitterkomplexe in the 'International Tables of Space-groups.' If these atoms are more numerous than the remaining atoms then the M.A. will not place the planes in order of decreasing density of atoms/unit area (*cf.* CaF<sub>2</sub>, above). Cuprite, Cu<sub>2</sub>O, illustrates this point well. The space-group is *Pn3m*. In this crystal the Cu atoms are arranged on a face-centred lattice (F) and the O atoms on a body-centred lattice (I). Accordingly, we obtain different M.A.'s depending on whether the criterion is the number of Cu+O, of Cu, or of O atoms per unit area:—

<i>Pn3m</i>	-	-	011	111	001	112	122, etc.
F(Cu)	-	-	111	001	011	113	133
I(O)	-	-	011	001	112	013	111

(d) If the structure consists of two or more sets of atoms of different kinds arranged in the same way but the sets being related to one another by a simple translation which does not introduce additional halvings, then the M.A. will correspond to the density of packing of atoms of one kind only. Thus the M.A. for NaCl relates to the density of packing of Na or Cl, but not of Na+Cl.

(4) *The Significance of the Experimental M.A.*—We shall not discuss the experimental M.A. in any detail since it would be difficult

to find any theoretical justification for it. The face developments of natural crystals grown from solution may have been influenced by a number of factors which are not taken into account in deriving the theoretical M.A.: (i.) the crystals have grown from solutions or melts of unknown composition, and it is well known that some impurities affect crystal habit, or (ii.) the observed development may represent an intermediate phase in the growth (or dissolution) of a crystal of type (i.), the conditions having changed during growth. (Variations in shape due to rapid crystallization can probably be ignored in the case of minerals.) The fact that it is possible to draw up an "observed M.A." for NaCl, for which the equilibrium shape is certainly the cube, shows that either or both of these factors are involved. The significance of the relative positions of the faces in such a series is therefore not known.

### *The Crystal Habits of Certain Halides.*

We shall now review, in the light of the ideas outlined above, the habits of certain halides, taking these as examples of solutes on the habits of which solute-solvent interaction has in general little effect.

#### *(a) Halides with the Rock-salt Structure.*

We have seen that (1) the most densely populated plane is the cube face (100), which contains equal numbers of positive and negative ions, and (2) the octahedron face (111) differs from the more closely packed (100) and (110) planes in that it contains ions all of the same type. We should therefore expect that in the absence of any complicating factors crystals with this structure will grow as cubes, but that any strong interaction between either the positive or negative ions and the solvent or added impurity will lead to the development of (111) faces.

All the halides of lithium, sodium, potassium and rubidium, and also CsF and  $\text{NH}_4\text{I}$ , crystallize with this structure. The chlorides, bromides and iodides crystallize as cubes from aqueous solution, melt or vapour. According to the literature, additions of other solutes cause habit modifications of which the following are typical:—NaCl, {111} favoured by NaOH,  $\text{H}_3\text{BO}_3$ ,  $\text{CaCl}_2$  and  $\text{FeCl}_3$ : {110} by  $\text{HgCl}_2$ . Certain organic compounds also favour the development of {111}, notably urea. Experiments on the growth of NaCl crystals (*e.g.* Spangenberg (1924), Schnorr (1928), Neuhaus (1928)) showed that various faces develop on a sphere of NaCl during growth, and that the rates of growth normal to the different faces are:  $v_{100} < v_{111} < v_{210} < v_{110}$ . The final product is a cube.

Data on the fluorides are less complete and far less satisfactory. Just as the chemistry of fluorine is very different from that of the other halogens owing to the extreme electronegativity of the smallest halogen, so there are some striking differences between fluorides and the other halides. Of particular interest here are the following. Although NaCl, NaBr and NaI crystallize from water as well-defined cubes, NaF



separates as ill-defined crystals with poorly-developed faces. Again, when some of the sparingly soluble fluorides such as  $\text{CaF}_2$  or  $\text{MgF}_2$  are precipitated from aqueous solution, they have a great tendency to form colloidal suspensions or gels. This is also true of some complex fluorides, *e.g.* silicofluorides and titanifluorides. These facts would appear to be related to the strong tendency of fluorine to form hydrogen bonds, and indicate greater interaction between the fluorides and water than between the other halides and this solvent.

An elaborate study has been made by Frondell (1940) of the effects of some 143 "cosolutes" on the habit of NaF crystallizing from aqueous solution. He gives few experimental details, and some of his assertions seem open to question. For example, he states that NaF crystallizes from water as cubes at all temperatures from 22–100° C. It is well known that NaF, unlike the other sodium halides, hydrolyzes ( $\text{H}_2\text{O} + 2\text{F}^- \rightarrow [\text{HF}_2]^- + \text{OH}^-$ ) so that it is difficult—if not impossible—to obtain a neutral solution of this salt. It may easily be verified that octahedral crystals of NaF separate from a boiled saturated solution which is left to stand for some days at room temperature. Such a solution is, however, distinctly alkaline. There is no doubt that the effect of  $\text{OH}^-$  ions on NaF is much greater than on the other sodium halides, but the statement of Frondell that "the habit-varying effect of the  $\text{OH}^-$  ion is peculiar to the fluorides NaF, LiF, and KF (RbF and CsF were not studied)" can easily be disproved. Frondell wished to show that the effect of  $\text{OH}^-$  ions is due to *substitutional* adsorption of  $\text{OH}^-$  for  $\text{F}^-$  (these ions having approximately the same radius) in the surface of the growing crystal. The fact that  $\text{OH}^-$  ions also lead to the development of (111) faces on NaCl and KBr obviously does nothing to strengthen such a theory. Although there are not yet sufficient data to show that the interaction of  $\text{F}^-$  ions with the solvent ( $\text{H}_2\text{O}$ ) has a decisive effect on the habits of fluorides, nevertheless the above facts together with the cubic habit of  $\text{CaF}_2$  do suggest that such interaction is much more pronounced than in the case of the other halides.

(b) *Halides with the Cæsium Chloride Structure.*

This structure consists of two interpenetrating simple cubic lattices, each ion being surrounded by eight of opposite charge, situated at the corners of a cube. The compositions of the faces are set out in Table VI. We should expect crystals with this structure to develop (110) faces (dodecahedron). The next most stable face should be (211), for although (100) is more densely populated than (211) it contains ions all of the same sign. The presence of ions or molecules which are preferentially adsorbed on a face consisting of all positive or all negative ions should accordingly encourage the development of cube faces. Groth records the habits of those cæsium and ammonium halides with this structure as follows:—

*CsCl*.—from aqueous solution {100} or {110}, CsBr and CsI, {110}



Without further data on CsCl it is not possible to say more than that there is general agreement with expectation.

$NH_4Cl$  and  $NH_4Br$ .—(Ammonium fluoride has the wurtzite ( $ZnS$ ) structure and  $NH_4I$  the rock-salt structure at ordinary temperatures).

$NH_4Cl$ .—On natural crystals  $\{110\}$  and  $\{211\}$ . The form  $\{211\}$  is commonly the only one on crystals grown from solution, forming distorted icositetrahedra. From aqueous solution containing  $FeCl_3$  this is the only form developed, but when crystallized from pure aqueous solution skeletal growths are normally obtained.

$NH_4Br$ .—From aqueous solution  $\{211\}$ . From solutions containing urea,  $CrCl_3$  or  $PbBr_2$ ,  $\{100\}$ .

The peculiar habits of  $NH_4Cl$  and  $NH_4Br$  are presumably connected with the tetrahedral shape of the ammonium ion. At higher temperatures both these salts, like the iodide at ordinary temperatures, have the rock-salt structure, in which the  $NH_4^+$  ion has attained spherical symmetry by free rotation.

TABLE VI.

$hkl$	Ions in Face	Total Number of Ions in Arbitrary Unit Area	Environments of Surface Ions
110	$M+X$	3.46	6
100	all $M$ or $X$	2.45	4
211	$M+X$	2.00	5 (7)
130	$M+X$	1.55	4 (6)
111	all $M$ or $X$	1.42	4 (7) (7)

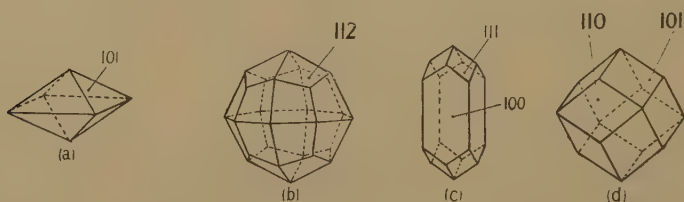
The presence of urea in an aqueous solution of  $NH_4Cl$  causes a change in habit from skeletal growths to compact cubes, and a similar effect is observed with  $NH_4Br$ . In the case of  $NaCl$ , which normally crystallizes as cubes, urea causes the development of octahedral faces. The effect is smaller with  $KCl$  and, according to Retgers, urea has no effect on the habits of  $KBr$  or  $KI$ . These facts indicate that although it is true that urea favours the development of those planes which contain ions all with the same charge, any explanation based solely on the electrical nature of the faces is not adequate. Many attempts have been made to interpret changes in habit due to foreign ions or molecules—see, for example, the work of Bunn (1933) and Royer (1934) on the relation between habit modification, mixed crystal formation and oriented overgrowths. It is equally doubtful if purely geometrical theories based on similarities in interatomic distances in certain planes of the crystalline solute and of the crystalline foreign material can be of general validity. In some cases (*e. g.* the effect of dyestuffs on alum,  $K_2SO_4$ , etc.) it is likely that the effects are due to interactions between particular atoms or groups of the

adsorbed molecule and atoms in the crystal face, *i.e.* that intra- as well as inter-molecular distances are important. In cases where comparatively large concentrations of habit-modifying material are present, it may well be necessary also to take into account the interactions between this material and the solute and solvent molecules in the solution.

(c) *The Tetramethylammonium Halides.*

The chloride, bromide and iodide are isomorphous: tetragonal, space-group  $P4nmm$ ,  $c:a=0.7108, 0.7131$  and  $0.7223$  respectively. The chloride will not be further considered here. The bromide crystallizes as bipyramids  $\{101\}$  (fig. 5, *a*), but the iodide with  $\{100\}$  and  $\{111\}$  (fig. 5, *c*). These habits are related to the internal structures of the crystals in very much the same way as for the ammonium halides. It is seen that  $a \approx \sqrt{2}c$ , so that the crystals can be referred to pseudo-cubic axes, and the structure may be described as a deformed CsCl structure.

Fig. 5.



(*a*) Tetramethylammonium bromide, (*b*) equivalent cubic form, (*c*) tetramethylammonium iodide, (*d*) equivalent cubic form.

Pseudo-cubic indices are shown in larger figures.

When referred to the pseudo-cubic axes the faces have the indices shown below:—

		Tetragonal.	Pseudo-cubic.
Tetramethylammonium bromide	-	101	112
„ „ iodide	-	{ 100 111	110 101

*i. e.*, the bromide develops the same (112) faces as  $\text{NH}_4\text{Br}$  but only eight, forming a tetragonal bipyramid, instead of the 24 forming an icositetrahedron, while the iodide develops 8 (111) and 4 (100) which, when equally developed, form a dodecahedron—compare the dodecahedral development of caesium halides. It is interesting that while the iodide develops these (111) and (100) faces, which are the most densely populated planes containing equal numbers of positive and negative ions, the bromide develops the next most densely populated (101) planes, just as  $\text{NH}_4\text{Cl}$  commonly grows (211) instead of the expected (110).

(d) *Certain Halides with Close-packed Structures.*

It is convenient to consider together two groups of halides in which (*a*) the halogen atoms and (*b*) the halogen and certain of the metal atoms

are arranged in cubic close-packing. Although the chlorine ions in NaCl are in the positions of cubic close-packing, this structure cannot be regarded as a close-packed lattice of  $\text{Cl}^-$  ions with  $\text{Na}^+$  ions in the interstices—witness the Cl—Cl distances in

		LiCl	NaCl	KCl	RbCl
Cl—Cl	- -	3.64	3.98	4.44	4.65 A.

On the other hand, the approximate constancy of the I—I distances in the following halides shows that we are justified in regarding them as based on a close-packed halogen lattice :—

		CuI	$\text{SnI}_4$	$\text{GeI}_4$	$\text{NiI}_2$
I—I	- -	4.28	4.32	4.20	4.18 A.

( $\text{GeI}_4$  is isomorphous with  $\text{SnI}_4$ ;  $\text{NiI}_2$  crystallizes with a non-cubic structure which we shall not deal with here.)

Halides of the first group include

- (1) those with the zinc-blende structure, *e. g.* CuCl, CuBr and CuI,
- (2) molecular halides, *e.g.*  $\text{SnI}_4$ .

If we refer the cubic close-packed halogen lattice to the usual face-centred cubic unit cell of side  $a$ , the relative densities of atoms in the simpler planes are as follows :—

$hkl$ - - -	111	100	110	113	133	210
Atoms/ $a^2$ - -	2.31	2.00	1.41	1.21	0.92	0.89

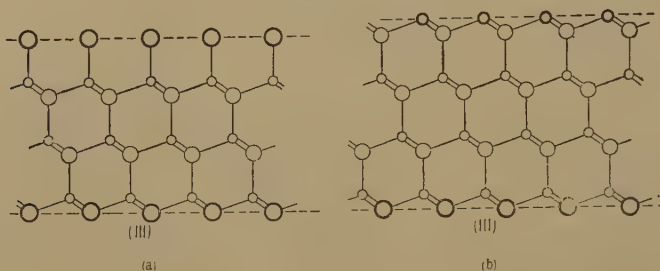
According to the literature CuCl forms tetrahedra from solution in hot hydrochloric acid and  $\{111\} + \{\bar{1}\bar{1}1\}$  when prepared by the action of sulphur dioxide on cupric chloride, and both CuBr and CuI crystallize as tetrahedra.  $\text{SnI}_4$  forms octahedra  $\{111\}$  with small  $\{210\}$  from solution in carbon disulphide.

Two points call for explanation—the development of tetrahedra instead of octahedra in the case of the cuprous halides, and the presence of  $\{210\}$  on  $\text{SnI}_4$ , this face being very much less closely packed with halogen atoms than the intermediate (100), (110), (113) and (133) planes. So far we have not taken into account the metal atoms. In a close-packed arrangement of atoms the largest holes are between six and four close-packed atoms respectively, and we shall refer to these as “octahedral” and “tetrahedral” holes. In CuI one-half of the tetrahedral holes are occupied. In fig. 6 are shown projections of the structure on  $(\bar{1}10)$  with (111) and  $(\bar{1}\bar{1}\bar{1})$  planes perpendicular to that of the paper. Since we are viewing the structure along the edge of a tetrahedral  $\text{CuI}_4$  or  $\text{ICu}_4$  co-ordination group, only three of the four nearest neighbours of each atom are visible. If we assume an outer layer of halogen atoms for both (111) and  $(\bar{1}\bar{1}\bar{1})$ , then in a (111) plane each of these atoms would be attached to three Cu atoms, instead of four as in the interior, but in the parallel  $(\bar{1}\bar{1}\bar{1})$  plane to only one Cu atom (fig. 6, *a*). Alternatively the  $(\bar{1}\bar{1}\bar{1})$  face could have an outer surface of

Cu atoms, each attached to three halogen atoms (fig. 6, *b*). In either case the structure and stability of (111) is different from that of ( $\bar{1}\bar{1}\bar{1}$ ) and the tetrahedral development results.

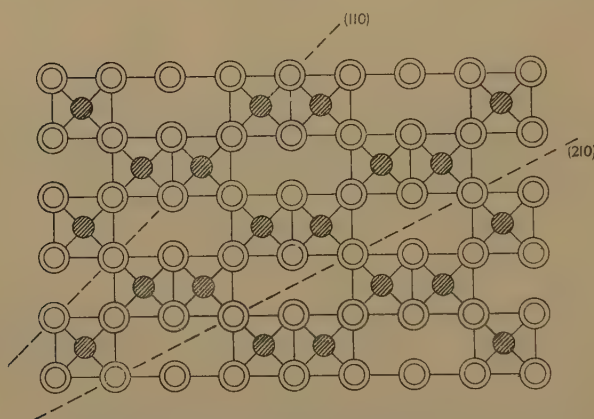
In  $\text{SnI}_4$  the arrangement of the metal atoms, in one-eighth of the tetrahedral holes, is such that there is no difference between parallel pairs of (111) faces. In a crystal consisting of discrete molecules we must, however, in estimating the relative stabilities of faces take into

Fig. 6.



Projections of the structure of CuI on (110) showing possible contours of (111) and ( $\bar{1}\bar{1}\bar{1}$ ) faces (see text).

Fig. 7.



Projection of the structure of  $\text{SnI}_4$  on (001).  
Shaded circles represent Sn atoms.

account the relation of the various faces to the molecules. The projection of the structure of  $\text{SnI}_4$  on (001) is shown in fig. 7, where it is seen that the trace of the (110) plane, for example, cuts through many molecules, so that a (110) face would have a very irregular contour. This is not so for the (100) or (210) planes. In other words, the relative densities of molecules in the various possible faces are different from those of halogen atoms. As regards the numbers of molecules of  $\text{SnI}_4$  per unit area, the most densely populated planes in this structure are (111), (100) and (210), with respectively 2.31, 2.00 and 1.79 molecules/ $a^2$ , where  $a$  is the



side of the unit cell. These figures should be compared with those given above. At present it is not possible to predict whether a combination of  $\{111\}$  and  $\{100\}$  or of  $\{111\}$  and  $\{210\}$  would be more stable.

The characteristic feature common to a number of complex halides  $A_xB_yX_z$  is that the larger positive ions A together with the halogen ions X form a close-packed array with the smaller positive ions B in the interstices. In such halides the ratio  $z:x=3$ . Examples are  $KMgF_3$ ,  $K_2PtCl_6$  and  $Cs_3Ti_2Cl_9$ . Little is known of the habit of  $KMgF_3$ , and  $Cs_3Ti_2Cl_9$  is rhombohedral and therefore cannot be directly compared with the cubic halides under discussion. The compositions of the three most densely populated planes are shown below. Accordingly it is found that  $K_2PtCl_6$ ,  $K_2SiF_6$ ,  $Rb_2SiF_6$ , etc., all with the same structure, crystallize from water as octahedra.

$hkl$	Proportions of A and X in Face	Total number of Atoms per Unit Area
111	A : 3X	2.31
100	A : X	2.00
110	A : X	1.41

#### SUMMARY.

The recent work of Donnay and Harker and others on the relation between the face developments and space-groups of crystals is reviewed. The morphologies of anthracene, diamond, zinc blende, rock-salt and fluorite are considered in some detail. Apart from the fact that morphological relationships which do not take account of the solvent are necessarily only of limited validity, it is shown that there are other complications, of a purely crystallographic nature, which severely restrict the usefulness of the Donnay-Harker generalization. The crystal habits of some metallic halides are discussed in relation to their crystal structures.

#### References.

- Bunn, Proc. Roy. Soc. cxli. A, p. 567 (1933).  
 Donnay, Amer. Min. xxiii. p. 5, *ibid.* xxiv. p. 184 (1939).  
 Donnay, Trans. Roy. Soc. Canada, xxxvi. sec. 4, p. 37 (1942).  
 Donnay, *ibid.* xxxvii. sec. 4, p. 43 (1943).  
 Donnay and Faessler, Univ. Toronto Studies, Geol. Ser. xlv. p. 19 (1941).  
 Donnay and Harker, *Compt. rend. Acad. Sci. Paris*, cciv. p. 274 (1937).  
 Donnay and Harker, Amer. Min. xxii. p. 446 (1937).  
 Donnay and Harker, *Naturaliste canadien*, lxvii. p. 33 (1940).  
 Friedel, *Compt. rend.* cxxxix. pp. 221, 314 (1904).  
 Friedel, Bull. Soc. franc. Min. xxviii. p. 6 (1905).  
 Friedel, *ibid.* xxx. p. 326 (1907).  
 Frondell, Amer. Min. xxv. p. 339 (1940).  
 Neuhaus, *Z. Krist.* lxviii. p. 15 (1928).  
 Peacock, Trans. Roy. Soc. Canada, xxxvi. sec. 4, p. 107 (1942).

- Peacock and Ferguson, Univ. Toronto Studies, Geol. Ser. xlviii. p. 65 (1943).  
 Robertson, Proc. Roy. Soc. exl. A, p. 79 (1933).  
 Robertson, *ibid.* cl. A, p. 106 (1935).  
 Royer, *Compt. rend.* cxcviii. pp. 185, 585 (1934).  
 Schnorr, *Z. Krist.* lxxviii. p. 1 (1928).  
 Shappell, Amer. Min. xxi. p. 75 (1936).  
 Spangenberg, *Z. Krist.* lix. p. 375 (1924).  
 Stranski, *Z. phys. Chem.* xvii. B, p. 127 (1932).  
 Taylor, Amer. Min. xxv. p. 327 (1940).  
 Tolansky, Proc. Roy. Soc. clxxxiv. A, pp. 41, 51 (1945).  
 Tremblay, J. Wash. Acad. Sci. xxxii. p. 327 (1942).

Imperial Chemical Industries Ltd.,  
 Research Laboratories,  
 Blackley, Manchester 9.

## XXVI. *Four-figure Tables of the Airy Function in the Complex Plane.*

By P. M. WOODWARD, B.A., and Mrs. A. M. WOODWARD, M.A.;  
 with the assistance of Miss R. HENSMAN, B.Sc.,  
 H. DAVIES, B.A., and Miss N. GAMBLE.

[Received May 30, 1946.]

### PREFATORY REMARKS.

THE immediate need for tables of the Airy function in the complex plane has arisen in connection with theoretical work on the propagation of electromagnetic waves through the earth's atmosphere, and it was for this particular purpose that the present tables were computed and produced as a report at the Telecommunications Research Establishment in February, 1945. It is not claimed that the tables are comprehensive, but they do provide material which, so far as is known at present, is not available elsewhere. The introductory matter, moreover, should be of interest independently of the tables, as it provides general suggestions with regard to interpolation in functions of a complex variable. Much of the labour associated with ordinary second and fourth difference bivariate interpolation may be avoided if suitable use be made of the Cauchy-Riemann equations and if the differences tabulated be modified accordingly.

### INTRODUCTION TO THE TABLES.

#### 1. *Definition of Functions Tabulated.*

The functions  $Ai(z)$  and  $Bi(z)$  are independent solutions of

$$d^2w/dz^2 = zw$$

such that

$$Ai(z)Bi'(z) - Ai'(z)Bi(z) = 1/\pi.$$

\* Communicated by R. A. Smith.

The power series

$$w_1 = 1 + \frac{1}{3!}z^3 + \frac{1 \cdot 4}{6!}z^6 + \frac{1 \cdot 4 \cdot 7}{9!}z^9 + \dots$$

$$w_2 = z + \frac{2}{4!}z^4 + \frac{2 \cdot 5}{7!}z^7 + \frac{2 \cdot 5 \cdot 8}{10!}z^{10} + \dots$$

were used for calculation of the functions by means of the relations

$$Ai(z) = pw_1 - qw_2,$$

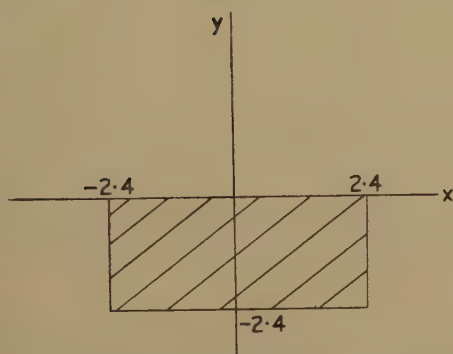
$$Bi(z) = 3^{1/2}(pw_1 + qw_2),$$

where

$$1/p = \left(-\frac{1}{3}\right)! 3^{2/3},$$

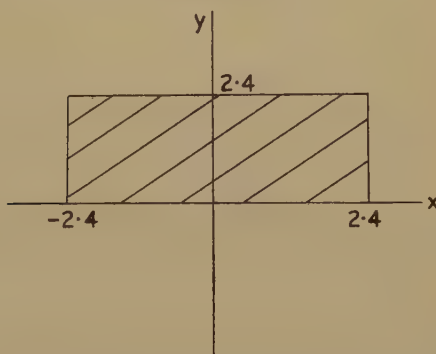
$$1/q = \left(-\frac{2}{3}\right)! 3^{1/3}.$$

Fig. 1.



Region of complex plane covered by tables.

Fig. 2.



Region covered by taking complex conjugates of tables.

## 2. Contents of Tables.

The following tables, each to four decimal places, cover the region

$$x = -2.4(0.2)2.4,$$

$$-y = 0.0(0.2)2.4,$$

in the lower half of the complex plane (see fig. 1) :—

TABLE	I.	..	..	Real part of $Ai(x+iy)$ .
TABLE	II.	..	..	Imaginary part of $Ai(x+iy)$ .
TABLE	III.	..	..	Real part of $Bi(x+iy)$ .
TABLE	IV.	..	..	Imaginary part of $Bi(x+iy)$ .
TABLE	V.	..	..	Real part of $Ai'(x+iy)$ .
TABLE	VI.	..	..	Imaginary part of $Ai'(x+iy)$ .
TABLE	VII.	..	..	Real part of $Bi'(x+iy)$ .
TABLE	VIII.	..	..	Imaginary part of $Bi'(x+iy)$ .

TABLE I.—Real part of  $Ai(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$\frac{x}{y}$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.3550 0 9	0.3037 24 8	0.2547 41 7	0.2098 50 5	0.1698 54 5	0.1353 54 2	0.1061 51 2	0.0830 46 1	0.0625 40 0	0.0470 34 0	0.0349 28 -1	0.0256 22 -1	0.0186 18 -1
-0.2	0.3550 4 9	0.3025 28 8	0.2527 44 7	0.2073 53 6	0.1671 56 4	0.1326 56 3	0.1036 52 1	0.0797 46 1	0.0605 40 0	0.0453 34 0	0.0335 28 0	0.0245 22 -1	0.0177 17 -1
-0.4	0.3545 17 10	0.2983 40 9	0.2461 55 8	0.1994 62 6	0.1587 63 4	0.1243 60 3	0.0958 54 1	0.0728 48 0	0.0545 40 0	0.0403 33 0	0.0294 27 -1	0.0211 21 -1	0.0150 16 -1
-0.6	0.3521 40 12	0.2901 62 10	0.2340 73 8	0.1852 76 6	0.1440 73 4	0.1100 67 2	0.0826 58 1	0.0611 49 0	0.0445 40 0	0.0319 32 -1	0.0225 25 -1	0.0157 19 -1	0.0108 14 -1
-0.8	0.3457 76 16	0.2755 93 12	0.2146 99 9	0.1635 96 5	0.1219 87 3	0.0890 75 1	0.0636 63 0	0.0445 51 -1	0.0304 39 -2	0.0203 22 -2	0.0084 16 -1	0.0017 11 -1	0.0051 11 -1
-1.0	0.3315 127 19	0.2516 137 14	0.1853 133 9	0.1322 121 5	0.0911 104 1	0.0605 85 0	0.0383 67 -2	0.0228 51 -2	0.0124 37 -3	0.0057 26 -3	0.0017 18 -2	-0.0006 11 -2	-0.0017 7 -1
-1.2	0.3045 197 23	0.2138 194 15	0.1425 176 8	0.0887 150 3	0.0500 122 -1	0.0235 94 -4	0.0063 69 -4	0.0039 48 -5	0.0093 32 -4	0.0114 20 -3	-0.0116 11 -3	-0.0106 5 -2	-0.0091 1 -1
-1.4	0.2575 291 27	0.1565 266 14	0.0820 227 6	0.0302 183 -1	-0.0033 139 -5	-0.0229 99 -7	-0.0325 67 -8	-0.0354 41 -7	-0.0341 23 -6	-0.0305 10 -5	-0.0259 1 -3	-0.0211 -4 -2	-0.0167 -6 -2
-1.6	0.1813 410 28	0.0724 353 12	-0.0013 284 0	-0.0466 214 -7	-0.0705 151 -11	-0.0792 98 -12	-0.0780 57 -11	-0.0710 28 -10	-0.0612 8 -8	-0.0505 -5 -5	-0.0404 -12 -4	-0.0313 -14 -3	-0.0237 -15 -1
-1.8	0.0638 557 25	-0.0471 451 4	-0.1129 340 -9	-0.1447 237 -17	-0.1525 152 -19	-0.1451 85 -19	-0.1291 36 -16	-0.1093 4 -13	-0.0889 -15 -10	-0.0700 -25 -7	-0.0536 -29 -4	-0.0400 -28 -4	-0.0292 -25 -2
-2.0	-0.1096 729 15	-0.2117 552 -11	-0.2584 386 -25	-0.2663 244 -32	-0.2496 133 -31	-0.2194 53 -27	-0.1837 -1 -22	-0.1478 -32 -16	-0.1150 -48 -11	-0.0869 -52 -7	-0.0640 -49 -3	-0.0459 -44 -1	-0.0322 -36 0
-2.2	-0.3561 915 -9	-0.4314 642 -38	-0.4423 405 -51	-0.4121 219 -47	-0.3598 83 -47	-0.2987 -7 -37	-0.2380 -59 -27	-0.1830 -84 -19	-0.1363 -90 -11	-0.0986 -85 -6	-0.0693 -74 -2	-0.0474 -60 0	-0.0316 -47 2
-2.4	-0.6939 1090 -53	-0.7150 692 -81	-0.6663 374 -86	-0.5794 141 -79	-0.4777 -14 -64	-0.3770 -104 -48	-0.2862 -146 -33	-0.2097 -155 -20	-0.1485 -145 10	-0.1017 -124 -3	-0.0673 -100 1	-0.0429 -77 3	-0.0263 -57 4



TABLE I. (continued).

$y \backslash x$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.3550 0 9	0.4063 -33 7	0.4542 -73 6	0.4948 -119 3	0.5236 -168 -2	0.5356 -214 -8	0.5262 -253 -16	0.4917 -275 -23	0.4299 -275 -30	0.3408 -245 -34	0.2274 -182 -34	0.0961 -85 -29	-0.0433 42 -19
-0.2	0.3550 4 9	0.4079 -29 9	0.4578 -70 7	0.5008 -117 3	0.5320 -168 -1	0.5463 -218 -8	0.5389 -260 -15	0.5056 -287 -23	0.4437 -290 -31	0.3532 -263 -36	0.2366 -199 -37	0.1005 -100 -32	-0.0453 32 -21
-0.4	0.3545 17 10	0.4123 -16 10	0.4683 -60 9	0.5184 -112 6	0.5572 -171 1	0.5789 -230 -6	0.5777 -283 -15	0.5483 -322 -25	0.4869 -336 -34	0.3921 -316 -42	0.2661 -254 -45	0.1151 -148 -42	-0.0504 3 -31
-0.6	0.3521 40 12	0.4182 7 14	0.4847 -40 13	0.5472 -100 10	0.5995 -171 6	0.6346 -247 -2	0.6451 -321 -13	0.6235 -382 -26	0.5639 -417 -40	0.4630 -412 -52	0.3214 -355 -60	0.1448 -239 -61	-0.0552 -63 -50
-0.8	0.3457 76 16	0.4233 43 18	0.5051 -8 13	0.5859 -78 19	0.6588 -166 13	0.7151 -267 4	0.7446 -372 -9	0.7371 -468 -27	0.6830 -537 -47	0.5755 -560 -67	0.4127 -516 -82	0.1989 -391 -89	-0.0533 -178 -83
-1.0	0.3315 127 19	0.4239 98 24	0.5260 45 29	0.6322 -37 29	0.7346 -146 25	0.8221 -281 16	0.8814 -431 -1	0.8977 -580 -24	0.8561 -704 -53	0.7446 -775 -85	0.5563 -761 -114	0.2929 -634 -132	-0.0329 -377 -132
-1.2	0.3045 197 23	0.4146 177 32	0.5422 125 39	0.6820 34 44	0.8248 -101 44	0.9571 -279 34	1.0613 -490 15	1.1164 -715 -15	1.1001 -924 -58	0.9918 -1075 -106	0.7770 -1120 -155	0.4514 -1011 -193	0.0263 -711 -208
-1.4	0.2575 291 27	0.3873 288 40	0.5456 245 54	0.7280 149 70	0.9247 -12 52	1.1197 -241 64	1.2901 -533 43	1.4068 -866 4	1.4370 -1201 -54	1.3475 -1482 -127	1.1109 -1635 -206	0.7126 -1582 -278	0.1583 -1255 -320
-1.6	0.1813 410 28	0.3309 439 48	0.5241 419 71	0.7586 329 92	1.0252 148 106	1.3058 -138 108	1.5717 -531 89	1.7838 -1010 43	1.8944 -1530 -2014	1.8523 -2014 -140	1.6100 -2356 -265	1.1343 -2434 -389	0.4185 -2124 -486
-1.8	0.0638 557 25	0.2302 638 54	0.4600 664 88	0.7554 602 125	1.1100 416 156	1.5049 74 173	1.9057 -438 163	2.2614 -1110 114	2.5052 -1891 16	2.5597 -2684 -132	2.3470 -3342 -322	1.8027 -3677 -530	0.8951 -3483 -714
-2.0	-0.1096 729 15	0.0652 890 53	0.3287 997 104	0.6910 1000 162	1.1519 840 221	1.6951 436 264	2.2821 -179 276	2.8490 -1091 234	3.3048 -2230 120	3.5366 -3483 -80	3.4208 -4649 -359	2.8431 -5452 -692	1.7260 -5564 -1022
-2.2	-0.3561 915 -9	-0.1892 1195 40	0.0968 1433 110	0.5253 1560 200	1.1080 1486 298	1.8369 1116 388	2.6741 360 442	3.5437 -832 429	4.3266 -2443 309	4.8625 -4354 58	4.9626 -6310 -337	4.4345 -7919 -851	3.1217 -8673 -1414
-2.4	-0.6939 1090 -53	-0.5634 1537 4	-0.2793 1978 97	0.2019 2318 225	0.9130 2431 382	1.8640 2160 545	3.0265 1348 678	4.3181 -136 728	5.5901 -2310 637	6.6232 -5154 343	7.1381 -8297 -189	6.8249 -1.3201 -952	5.3965 -1.3201 -1870

TABLE II.—Imaginary part of  $\text{Ai}(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x \backslash y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0	0.0000 0 0 0
-0.2	0.0522 -28 0	0.0508 -20 1	0.0474 -13 1	0.0427 -6 2	0.0373 -1 1	0.0318 2 1	0.0265 4 1	0.0216 6 1	0.0173 6 1	0.0136 6 1	0.0105 6 0	0.0080 5 0	0.0060 4 0
-0.4	0.1073 -57 1	0.1036 -39 2	0.0960 -24 3	0.0860 -11 3	0.0747 -2 3	0.0634 5 2	0.0525 10 2	0.0426 12 2	0.0340 13 1	0.0266 13 0	0.0205 12 0	0.0155 10 0	0.0116 9 0
-0.6	0.1681 -85 1	0.1604 -57 2	0.1470 -33 4	0.1303 -13 5	0.1123 1 4	0.0944 11 3	0.0775 17 2	0.0624 20 2	0.0493 21 1	0.0383 20 1	0.0292 18 0	0.0220 16 0	0.0163 13 0
-0.8	0.2372 -111 4	0.2228 -70 6	0.2013 -37 8	0.1760 -10 7	0.1497 9 7	0.1242 21 6	0.1008 28 4	0.0802 31 3	0.0626 30 2	0.0480 30 1	0.0362 25 0	0.0269 21 0	0.0197 17 0
-1.0	0.3174 -133 8	0.2922 -77 11	0.2591 -33 12	0.2226 1 11	0.1861 23 9	0.1519 37 7	0.1212 43 5	0.0948 44 3	0.0728 42 2	0.0549 37 1	0.0407 32 0	0.0297 26 0	0.0214 21 0
-1.2	0.4108 -146 15	0.3692 -73 18	0.3201 -17 17	0.2691 22 15	0.2202 47 12	0.1758 59 9	0.1373 63 6	0.1051 61 4	0.0788 55 2	0.0581 47 1	0.0420 39 1	0.0299 31 -1	0.0209 25 -1
-1.4	0.5187 -144 26	0.4534 -51 39	0.3826 15 24	0.3132 58 20	0.2495 82 15	0.1938 90 10	0.1471 89 6	0.1092 81 3	0.0793 70 1	0.0565 58 -1	0.0394 46 -2	0.0270 36 -2	0.0181 27 -2
-1.6	0.6408 -116 41	0.5425 -3 54	0.4435 72 33	0.3514 114 25	0.2705 132 17	0.2027 132 11	0.1479 121 5	0.1052 104 2	0.0729 86 -1	0.0491 68 -2	0.0322 52 -3	0.0205 38 -3	0.0125 27 -3
-1.8	0.7742 -46 63	0.6315 84 63	0.4969 161 42	0.3779 195 30	0.2782 199 19	0.1983 184 10	0.1366 159 4	0.0907 129 -1	0.0578 101 -3	0.0350 76 -4	0.0199 54 -5	0.0102 38 -4	0.0043 25 -4
-2.0	0.9116 88 91	0.7117 226 73	0.5338 292 52	0.3847 305 34	0.2659 285 18	0.1754 246 7	0.1094 199 -1	0.0633 154 -5	0.0327 113 -7	0.0134 79 -8	0.0021 53 -7	-0.0038 33 -6	-0.0064 20 -5
-2.2	1.0395 313 127	0.7686 441 93	0.5411 476 61	0.3607 449 34	0.2249 389 13	0.1279 314 0	0.0623 239 -9	0.0206 173 -12	-0.0037 118 -13	-0.0161 75 -12	-0.0209 44 -10	-0.0211 23 -8	-0.0190 9 -5
-2.4	1.1349 666 167	0.7807 749 111	0.5004 720 64	0.2915 626 27	0.1450 505 2	0.0490 381 -14	-0.0087 270 -20	-0.0393 179 -22	-0.0517 110 -20	-0.0530 59 -17	-0.0482 26 -13	-0.0407 5 -9	-0.0326 -7 -6

TABLE II. (continued).

$\frac{y}{x}$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.2	0.0522	0.0508	0.0457	0.0363	0.0220	0.0028	-0.0209	-0.0478	-0.0759	-0.1028	-0.1250	-0.1391	-0.1419
-0.4	-0.28	-0.37	-0.44	-0.49	-0.50	-0.45	-0.33	-0.14	13	46	81	114	140
-0.6	0.1073	0.1053	0.0959	0.0775	0.0490	0.0101	-0.0384	-0.0941	-0.1531	-0.2101	-0.2581	-0.2898	-0.2979
-0.8	0.1681	0.1673	0.1551	0.1289	0.0865	0.0272	-0.0483	-0.1367	-0.2322	-0.3262	-0.4077	-0.4642	-0.4836
-1.0	-0.111	-0.155	-0.198	-0.235	-0.256	-0.253	-0.217	-0.140	-0.3124	-0.4545	-0.5821	-0.6762	-0.7174
-1.2	0.3174	0.3294	0.3217	0.2879	0.2221	0.1200	-0.0194	-0.1927	-0.3905	-0.5970	-0.7898	-0.9415	-1.0222
-1.4	-0.146	-0.234	-0.331	-0.426	-0.505	-0.546	-0.529	-0.431	-0.235	-0.065	-0.458	-0.908	-1.357
-1.6	0.5187	0.5695	0.5938	0.5780	0.5077	0.3696	0.1547	-0.1390	-0.5025	-0.9130	-1.3319	-1.7055	-2.19685
-1.8	-0.116	-0.270	-0.461	-0.677	-0.894	-0.83	-0.6621	-0.2177	-0.3995	-1.1686	-2.0345	-2.9035	-3.6444
-2.0	0.7742	0.9117	1.0248	1.0880	1.0704	0.9388	0.6621	0.2177	-0.1750	-1.002	-0.63	-0.287	-0.77
-2.2	-0.46	-0.239	-0.495	-0.805	-1.142	-0.85	-0.170	-0.263	-0.1445	-1.1708	-0.824	-0.3876	-0.8830
-2.4	0.9116	1.1195	1.3124	1.4571	1.5110	1.4242	1.1454	0.6307	-0.2632	-1.987	-0.824	-0.556	-0.299
-2.6	0.103	0.103	0.100	0.092	0.085	0.075	0.068	0.060	0.053	0.046	0.039	0.032	0.025
-2.8	1.0395	1.3406	1.6463	1.9158	2.0919	2.1027	1.8679	1.3697	0.3697	-0.9693	-2.6590	-4.5641	-6.4497
-3.0	313	59	-349	-922	-1644	-2457	-3250	-3852	-4044	-3581	-2240	-113	3444
-3.2	127	155	168	153	95	-16	-186	-407	-657	-886	-1027	-998	-716
-3.4	1.1349	1.5544	-2.0138	2.4655	2.8364	3.0270	2.9169	2.3772	1.2937	-0.4023	-2.6947	-5.4437	-8.3548
-3.6	666	417	-54	-785	-1784	-3000	-4305	-5476	-6194	-6069	-4697	-1763	2839
-3.8	167	222	263	272	224	97	-126	-447	-842	-1253	-1581	-1696	-1451

TABLE III.—Real Part of  $Bi(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x \backslash y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.6149 0 -14 0.6150 -7 -14 0.6159 -28 -11 0.6196 -59 -7 0.6293 -98 -1 0.6489 -138 6 0.6821 -172 16 0.7325 -189 29 0.8015 -177 46 0.8878 -118 69 0.9854 10 97 1.0812 236 132 1.1522 594 171	0.7055 56 -15 0.7027 49 -14 0.6952 27 -11 0.6850 -5 -6 0.6753 -43 1 0.6700 -80 10 0.6725 -106 20 0.6854 -113 32 0.7094 -86 47 0.7415 -14 63 0.7746 81 122 0.7947 339 100 0.7800 656 116	0.8018 128 -19 0.7954 119 -17 0.7773 94 -12 0.7499 55 -6 0.7171 11 2 0.6831 -31 12 0.6520 -61 22 0.6269 -68 33 0.6082 -42 45 0.5934 28 55 0.5753 154 64 0.5413 343 70 0.4724 601 70	0.9111 219 -24 0.9002 207 -22 0.8869 173 -17 0.8705 94 -12 0.8205 122 -8 0.7599 63 2 0.6930 6 13 0.6253 -38 25 0.5612 -57 35 0.5025 -42 43 0.4477 15 49 0.3909 121 50 0.3216 276 46 0.2244 476 35	1.0424 333 -34 1.0259 317 -32 0.9779 269 -24 0.9033 197 -13 0.8090 113 1 0.7034 30 15 0.5947 -39 28 0.4896 -81 38 0.3923 -85 44 0.3031 -46 45 0.2182 38 41 0.1291 161 31 0.0237 315 11	1.2074 483 -49 1.1835 459 -45 1.1140 389 -35 1.0059 286 -19 0.8694 163 1 0.7167 39 17 0.5599 -68 33 0.4096 -142 45 0.2732 -172 50 0.1536 -153 47 0.0488 -87 39 -0.0475 16 22 -0.1457 141 -1	1.4211 682 -72 1.3873 647 -66 1.2894 546 -52 1.1373 394 -29 0.9462 213 3 0.7337 30 23 0.5180 -131 44 0.3151 -249 58 0.1365 -309 62 -0.0117 -308 58 -0.1295 -251 42 -0.2227 -151 22 -0.3009 -31 -7	1.7037 954 -106 1.6564 902 -97 1.5197 754 -75 1.3084 531 -42 1.0442 267 4 0.7535 33 62 0.4626 -239 85 0.1951 -414 81 -0.0317 -509 85 -0.2082 -521 76 -0.3333 -458 56 -0.4131 -340 25 -0.4591 -197 -8	2.0825 1333 -155 2.0165 1256 -144 1.8261 1038 -110 1.5328 711 -60 1.1690 324 4 0.7728 50 92 0.3829 -409 122 0.0832 -661 117 -0.2514 -796 85 -0.4575 -811 107 -0.5834 -731 76 -0.6378 -555 36 -0.6370 -354 -9	2.5959 1869 -230 2.5034 1756 -212 2.2371 1433 -161 1.8288 951 -86 1.3262 385 -1 0.7850 -183 78 0.2615 -674 141 -0.1958 -1027 175 -0.5518 -1206 179 -0.7888 -1210 155 -0.9061 -1061 106 -0.9182 -806 48 -0.8502 -503 -12	3.2981 2639 -342 3.1676 2470 -314 2.7927 1991 -236 2.2207 1279 -121 1.5218 447 6 0.7782 -380 127 0.0715 -1082 217 -0.5288 -1570 265 -0.9744 -1798 266 -1.2423 -18740 223 -1.3357 -1508 150 -1.2796 -1104 73 -1.1136 -639 -25	4.2670 3755 -512 4.0814 3503 -469 3.5494 2788 -348 2.7415 1730 -173 1.7622 501 25 0.7325 -705 204 -0.2285 -1710 338 -1.0212 -2384 404 -1.5789 -2659 396 -1.8740 -2544 325 -1.9175 -2107 298 -1.7520 -1465 284 -2.3714 -1896 75 -1.8421 -748 -55	



TABLE III. (continued).

$\frac{x}{y}$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.6149	0.5245	0.4300	0.3288	0.2198	0.1040	-0.0158	-0.1347	-0.2460	-0.3414	-0.4123	-0.4504	-0.4491
	0	-42	-69	-79	-70	-42	8	75	157	246	330	396	431
-0.2	-14	-15	-17	-18	-21	-21	-19	-15	-6	4	18	32	45
	0.6150	0.5267	0.4335	0.3328	0.2234	0.1062	-0.0161	-0.1384	-0.2538	-0.3537	-0.4289	-0.4703	-0.4708
	-7	-50	-77	-88	-81	-52	-2	68	154	248	339	412	454
-0.4	-14	-15	-16	-19	-21	-22	-20	-16	-9	3	16	32	47
	0.6159	0.5339	0.4449	0.3458	0.2353	0.1137	-0.0161	-0.1488	-0.2770	-0.3908	-0.4794	-0.5318	-0.5383
	-28	-71	-102	-117	-112	-84	-32	44	142	252	364	461	524
-0.6	-11	-13	-16	-20	-23	-25	-26	-21	-14	-2	15	34	53
	0.6196	0.5484	0.4666	0.3707	0.2585	0.1299	-0.0126	-0.1635	-0.3142	-0.4532	-0.5665	-0.6395	-0.6586
	-59	-106	-143	-164	-166	-141	-87	-1	116	255	404	543	646
-0.8	-7	-10	-15	-20	-26	-30	-32	-31	-23	-10	10	34	62
	0.6293	0.5735	0.5027	0.4121	0.2984	0.1605	-0.0001	-0.1778	-0.3628	-0.5409	-0.6941	-0.8019	-0.8441
	-98	-151	-198	-232	-246	-229	-175	-77	66	248	453	659	831
-1.0	-1	-7	-12	-21	-30	-38	-44	-45	-39	-25	-1	33	72
	0.6489	0.6139	0.5587	0.4769	0.3632	0.2143	0.0302	-0.1841	-0.4177	-0.6532	-0.8669	-1.0303	-1.1132
	-138	-202	-266	-321	-355	-355	-307	-199	-24	214	501	808	1087
-1.2	6	0	-9	-21	-34	-48	-61	-67	-65	-50	-20	25	83
	0.6821	0.6745	0.6414	0.5741	0.4639	0.3040	0.0917	-0.1699	-0.4696	-0.7866	-1.0898	-1.3398	-1.4917
	-172	-253	-342	-430	-499	-530	-81	-99	-103	129	528	981	1427
-1.4	16	9	-3	-18	-39	-60	-81	-81	-103	-91	-56	4	88
	0.7325	0.7603	0.7585	0.7144	0.6147	0.4473	0.2040	-0.1160	-0.5026	-0.9320	-1.3650	-1.7473	-2.0137
	-189	-295	-422	-557	-681	-765	-775	-678	-442	-48	497	1157	1853
-1.6	29	22	9	-12	-39	-73	-109	-140	-159	-103	-117	-40	78
	0.8015	0.8754	0.9177	0.9106	0.8340	0.6676	0.3947	0.0069	-0.4900	-1.0714	-1.6890	-2.2702	-2.7216
	-177	-314	-491	-695	-901	-1072	-1159	-1109	-864	-384	345	1289	2354
-1.8	46	41	28	3	-34	-82	-139	-195	-238	-252	-220	-127	33
	0.8878	1.0216	1.1258	1.1762	1.1436	0.9957	0.7025	0.2423	-0.3891	-1.1703	-2.0456	-2.9210	-3.6652
	-118	-291	-532	-829	-1154	-1461	-1683	-1735	-1526	-975	-32	1288	2883
-2.0	69	68	57	30	-17	-84	-169	-262	-345	-396	-385	-285	-74
	0.9854	1.1964	1.3866	1.5245	1.5687	1.4707	1.1799	0.6534	-0.1337	-1.1684	-2.3957	-3.6981	-4.8965
	10	-199	-514	-930	-1422	-1932	-2374	-2623	-2537	-1967	-803	993	3328
-2.2	97	107	103	78	23	-67	-190	-336	-486	-602	-642	-554	-297
	1.0812	1.3901	1.6979	1.9651	2.1358	2.1394	1.8963	1.3296	0.3815	-0.9647	-2.6603	-4.5698	-6.4581
	236	1	-391	-952	-1664	-2467	-3252	-3848	-4035	-3568	-2225	130	3460
-2.4	132	159	171	154	97	-14	-185	-407	-655	-885	-1025	-996	-716
	1.1522	1.5825	2.0469	2.4996	2.8685	3.0550	2.9395	2.3939	1.3045	-0.3908	-2.6938	-5.4453	-8.3597
	594	-361	-96	-816	-1805	-3012	-4311	-5476	-6189	-6061	-4687	-1751	2850
	171	225	266	274	226	98	-125	-446	-841	-1251	-1581	-1695	-1449

TABLE IV.—Imaginary part of  $B(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x \backslash y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0
-0.2	-0.0888 -49 0	-0.0913 -64 2	-0.1001 -80 4	-0.1170 -100 8	-0.1438 -128 12	-0.1837 -168 18	-0.2404 -226 26	-0.3201 -312 40	-0.4312 -437 59	-0.5866 -623 88	-0.8049 -897 133	-1.1141 -1307 202	-1.5557 -1922 309
-0.4	-0.1728 -99 0	-0.1763 -125 3	-0.1923 -155 8	-0.2239 -193 13	-0.2750 -244 22	-0.3506 -318 33	-0.4584 -426 50	-0.6093 -584 73	-0.8191 -816 108	-1.1116 -1158 161	-1.5212 -1664 242	-2.0992 -2415 367	-2.9219 -3539 560
-0.6	-0.2468 -149 0	-0.2487 -184 3	-0.2690 -223 10	-0.3118 -272 18	-0.3818 -339 29	-0.4860 -436 44	-0.6342 -577 65	-0.8406 -785 95	-1.1264 -1089 141	-1.5221 -1535 208	-2.0731 -2192 309	-2.8458 -3163 466	-3.9387 -4606 706
-0.8	-0.3060 -201 0	-0.3027 -240 2	-0.3236 -281 10	-0.3726 -333 20	-0.4551 -404 33	-0.5782 -510 49	-0.7528 -664 71	-0.9943 -891 104	-1.3259 -1223 150	-1.7809 -1707 220	-2.4084 -325 289	-3.2801 -484 421	-4.5008 -4978 726
-1.0	-0.3450 -259 0	-0.3328 -295 2	-0.3500 -329 9	-0.4003 -373 20	-0.4881 -438 32	-0.6199 -535 48	-0.8055 -680 68	-1.0598 -895 97	-1.4043 -1208 138	-1.8708 -1661 199	-2.5051 -2315 289	-3.3733 -3262 421	-4.5712 -4634 620
-1.2	-0.3579 -327 0	-0.3333 -350 3	-0.3437 -368 7	-0.3909 -394 19	-0.4776 -438 30	-0.6085 -512 42	-0.7909 -628 58	-1.0366 -803 79	-1.3632 -1056 108	-1.7963 -1419 150	-2.3726 -1933 208	-3.1439 -2657 295	-4.1833 -3679 417
-1.4	-0.3380 -410 0	-0.2989 -408 6	-0.3006 -399 6	-0.3422 -397 16	-0.4236 -410 25	-0.5462 -448 33	-0.7138 -519 42	-0.9337 -632 52	-1.2173 -798 66	-1.5812 -1029 82	-2.0486 -1343 105	-2.6512 -1761 131	-3.4310 -2311 162
-1.6	-0.2769 -513 0	-0.2237 -472 6	-0.2176 -424 7	-0.2540 -382 16	-0.3288 -356 21	-0.4394 -351 24	-0.5853 -368 24	-0.7681 -410 23	-0.9922 -474 19	-1.2638 -557 11	-1.5912 -649 6	-1.9835 -735 37	-2.4489 -781 96
-1.8	-0.1643 -639 0	-0.1012 -542 11	-0.0923 -442 11	-0.1277 -353 18	-0.1986 -282 19	-0.2978 -230 16	-0.4201 -193 9	-0.5617 -165 5	-0.7198 -23 54	-0.8909 -54 5101	-1.0688 -101 629	-1.2419 -180 1072	-1.3879 -306 1795
-2.0	0.0124 -788 0	0.0755 -614 10	0.0771 -448 23	0.0337 -305 27	-0.0403 -187 23	-0.1333 -92 13	-0.2356 -10 3	-0.3388 -25 25	-0.4341 -56 189	-0.5101 -101 358	-0.5495 -169 629	-0.5246 -274 1072	-0.3901 -440 1795
-2.2	0.2681 -951 6	0.3135 -674 34	0.2911 -430 45	0.2254 -229 43	0.1365 -70 43	0.0403 58 14	-0.0502 172 -8	-0.1234 294 -37	-0.1669 455 -74	-0.1643 690 -126	-0.0917 1053 -479	0.0878 1619 -312	0.4319 2502 1.0077
-2.4	0.6188 -1106 50	0.6186 -699 75	0.5477 -366 79	0.4397 -110 67	0.3201 -80 47	0.2080 223 21	0.1181 346 -8	0.0629 476 -40	0.0555 647 -79	0.1136 898 -129	0.2625 1279 -196	0.5409 1859 -293	1.0077 2736 -434

TABLE IV. (continued).

$x$ $y$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.2	0.0888	0.0913	0.0972	0.1051	0.1130	0.1191	0.1213	0.1176	0.1062	0.0861	0.0571	0.0201	0.0222
-0.4	0.1728	0.1791	0.1925	0.2100	0.2278	0.2422	0.2486	0.2429	0.2214	0.1814	0.1222	0.0459	0.0427
-0.6	0.2468	0.2598	0.2838	0.3143	0.3462	0.3731	0.3882	0.3844	0.3553	0.2961	0.2051	0.0846	0.0584
-0.8	0.3060	0.3293	0.3683	0.4173	0.4693	0.5158	0.5467	0.5514	0.5197	0.4434	0.3185	0.1464	0.0635
-1.0	0.3450	0.3830	0.4424	0.5169	0.5978	0.6739	0.7315	0.7551	0.7293	0.6405	0.4801	0.2475	0.0469
-1.2	0.3579	0.4151	0.5011	0.6096	0.7307	0.8503	0.9502	1.0087	1.0027	0.9105	0.7157	0.4127	0.0107
-1.4	0.3380	0.4179	0.5368	0.6891	0.8643	1.0458	1.2099	1.3270	1.3633	1.2846	1.0622	0.6802	0.1427
-1.6	0.2769	0.3812	0.5381	0.7445	0.9906	1.2575	1.5159	1.7261	1.8397	1.8044	1.5719	1.1077	0.4040
-1.8	0.1643	0.2911	0.4887	0.7588	1.0943	1.4759	1.8687	2.2210	2.4654	2.5239	2.3176	1.7812	0.8822
-2.0	0.0124	0.1293	0.3653	0.7055	1.1494	1.6802	2.2592	2.8218	3.2768	3.5105	3.3986	2.8262	1.7150
-2.2	0.2681	0.1275	0.1360	0.5459	1.1139	1.8318	2.6614	3.5264	4.3075	4.8439	4.9462	4.4215	3.1127
-2.4	0.6188	0.5081	0.2414	0.2248	0.9237	1.8651	3.0208	4.3079	5.5777	6.6104	7.1263	6.8151	5.3893
	1106	1560	2004	2346	2458	2187	1372	114	2317	5140	1.1225	1.2251	1.3199
	50	7	99	227	384	545	678	729	637	343	188	952	1870

TABLE V.—Real Part of  $Ai'(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x \backslash y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	-0.2588 142	-0.2524 101	-0.2358 64	-0.2128 33	-0.1864 8	-0.1591 -10	-0.1328 -21	-0.1085 -28	-0.0870 -31	-0.0685 -31	-0.0531 -29	-0.0405 -25	-0.0304 -22
0.2	-0.2659 142	-0.2575 99	-0.2390 61	-0.2144 29	-0.1868 5	-0.1586 -13	-0.1317 -24	-0.1071 -30	-0.0855 -32	-0.0670 -31	-0.0517 -29	-0.0392 -26	-0.0294 -22
0.4	-0.2772 141	-0.2724 93	-0.2483 51	-0.2189 19	-0.1876 -6	-0.1568 -22	-0.1282 -32	-0.1027 -36	-0.0807 -36	-0.0623 -34	-0.0473 -31	-0.0354 -26	-0.0261 -22
0.6	-0.3226 136	-0.2966 79	-0.2626 33	-0.2252 1	-0.1878 -25	-0.1528 -38	-0.1215 -44	-0.0946 -45	-0.0723 -43	-0.0542 -38	-0.0399 -33	-0.0289 -27	-0.0206 -21
0.8	-0.3714 123	-0.3286 54	-0.2801 3	-0.2312 -32	-0.1855 -52	-0.1448 -62	-0.1103 -5	-0.0820 -59	-0.0595 -52	-0.0422 -43	-0.0293 -35	-0.0198 -27	-0.0130 -21
1.0	-0.4325 93	-0.3658 11	-0.2978 -45	-0.2340 -77	-0.1778 -92	-0.1306 -93	-0.0927 -8	-0.0634 -86	-0.0416 -75	-0.0259 -62	-0.0151 -36	-0.0079 -26	-0.0034 -18
1.2	-0.5027 36	-0.4040 -27	-0.3108 -39	-0.2289 -140	-0.1608 -144	-0.1070 -133	-0.0665 -114	-0.0374 -93	-0.0176 -71	-0.0048 -52	0.0027 -36	0.0065 -24	0.0080 -14
1.4	-0.5761 -65	-0.4360 -167	-0.3121 -216	-0.2095 -226	-0.1293 -211	-0.0701 -181	-0.0288 -145	-0.0021 -110	0.0136 -78	0.0214 -52	0.0240 -32	0.0233 -17	0.0209 -7
1.6	-0.6425 -89	-0.4509 -328	-0.2915 -356	-0.1674 -337	-0.0767 -291	-0.0150 -234	0.0233 -175	0.0441 5	0.0525 10	0.0528 10	0.0485 8	0.0418 6	0.0344 4
1.8	-0.6851 -125	-0.4325 -87	-0.2350 -542	-0.0913 -474	0.0052 -382	0.0634 -286	0.0930 16	0.1025 17	0.0993 24	0.0888 14	0.0751 10	0.0609 8	0.0477 6
2.0	-0.6779 -160	-0.3574 -99	-0.1237 -780	0.0325 -634	0.1253 -476	0.1704 -330	0.1824 -207	0.1734 30	0.1530 27	0.1278 18	0.1022 23	0.0789 35	0.0590 38
2.2	-0.5821 -1414	-0.1940 -1290	0.0659 -1064	0.2197 -804	0.2030 -557	0.3102 -348	0.2923 -186	0.2554 41	0.2110 33	0.1668 23	0.1269 15	0.0933 9	0.0665 4
2.4	-0.3434 -2147	-0.0992 -1803	0.3621 -1377	0.4373 -959	0.5160 -598	0.4843 -317	0.4203 -116	0.3440 72	0.2855 57	0.2010 40	0.1449 27	0.1008 15	0.0676 3



TABLE V. (continued).

$\frac{x}{y}$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	-0.2588 142	-0.2510 183	-0.2250 218	-0.1774 241	-0.1058 243	-0.0102 218	0.1070 159	0.2398 62	0.3785 -70	0.5100 -231	0.6183 -404	0.6862 -565	0.6980 -687
-0.2	-0.2659 142	-0.2602 185	-0.2360 224	-0.1895 251	-0.1181 257	-0.0212 236	0.0989 178	0.2365 81	0.3819 -56	0.5215 -224	0.6385 -409	0.7147 -585	0.7327 -723
-0.4	-0.2872 141	-0.2879 192	-0.2694 241	-0.2268 280	-0.1563 300	-0.0561 289	0.0726 238	0.2248 140	0.3906 -8	0.5553 -200	0.6996 -420	0.8019 -643	0.8403 -831
-0.6	-0.3226 136	-0.3348 201	-0.3270 269	-0.2923 330	-0.2248 373	-0.1203 383	0.0221 347	0.1987 250	0.3998 85	0.6088 -145	0.8028 -428	0.9538 -733	1.0316 -1016
-0.8	-0.3714 123	-0.4019 207	-0.4115 302	-0.3910 397	-0.3309 478	-0.2232 525	-0.0636 516	0.1470 429	0.3998 247	0.6764 -36	0.9486 -411	1.1791 -845	1.3252 -1280
-1.0	-0.4325 93	-0.4896 202	-0.5263 334	-0.5295 478	-0.4850 616	-0.3791 722	-0.2015 764	0.0516 705	0.3742 514	0.7468 168	1.1350 -331	1.4889 -951	1.7474 -1620
-1.2	-0.5027 36	-0.5973 173	-0.6743 352	-0.7158 563	-0.7009 785	-0.6077 984	-0.4166 102	-0.1154 133	0.2960 156	0.7992 531	1.3534 -124	1.8934 -1000	2.3321 -2016
-1.4	-0.5761 65	-0.7221 102	-0.8573 337	-0.9583 635	-0.9955 974	-0.9352 1313	-0.7441 1587	-0.3956 1711	0.1220 231	0.7964 250	1.5823 316	2.3966 -900	3.1184 -2413
-1.6	-0.6425 -230	-0.8564 -41	-1.0734 255	-1.2640 662	-1.3876 1162	-1.3946 1704	-1.2315 2207	-0.8490 2549	-0.2141 2585	0.6754 2162	1.7763 1156	2.9874 -486	4.1451 -2691
-1.8	-0.6851 -91	-0.9856 -109	-1.3141 -113	-1.6352 -95	-1.8954 -45	-2.0248 36	-1.9409 158	-1.5599 306	-0.8127 461	0.3333 590	1.8494 645	3.6219 578	5.4381 343
-2.0	-0.6779 -125	-1.0842 -161	-1.5593 -181	-2.0641 -176	-2.5321 -131	-2.8677 -27	-2.9496 138	-2.6430 361	-1.8203 616	-0.3931 860	1.6488 1022	4.1958 521	6.9855 771
-2.2	-0.5821 -160	-1.1101 -224	-1.7702 -278	-2.5248 -301	-3.2967 -268	-3.9615 -156	-4.3480 58	-4.2485 375	-3.4455 779	-1.7543 1208	0.9169 1566	4.4985 1717	8.6935 1515
-2.4	-0.3434 -2147	-0.9989 -2283	-1.8797 -2045	-2.9603 -1253	-4.1602 259	-5.3273 2580	-6.2300 5658	-6.5626 9218	-5.9738 918	-4.1225 2046	-0.7637 3268	4.1415 4303	10.3082 4741

TABLE VI.—Imaginary part of  $A'(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$\frac{x}{y}$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	0.0007	0.0128	0.0210	0.0256	0.0275	0.0273	0.0256	0.0230	0.0200	0.0169	0.0139	0.0112	0.0088
0.4	0.0036	0.0296	0.0454	0.0541	0.0572	0.0550	0.0521	0.0465	0.0401	0.0337	0.0276	0.0221	0.0173
0.6	0.0196	0.0548	0.0769	0.0882	0.0909	0.0876	0.0802	0.0707	0.0603	0.0501	0.0406	0.0322	0.0250
0.8	0.0480	0.0931	0.1195	0.1308	0.1308	0.1230	0.1105	0.0957	0.0803	0.0657	0.0524	0.0410	0.0314
1.0	0.0980	0.1501	0.1771	0.1847	0.1783	0.1631	0.1429	0.1211	0.0995	0.0798	0.0625	0.0479	0.0360
1.2	0.1782	0.2323	0.2542	0.2523	0.2346	0.2078	0.1770	0.1459	0.1169	0.0913	0.0697	0.0521	0.0382
1.4	0.2991	0.3469	0.3548	0.3355	0.2998	0.2563	0.2112	0.1685	0.1307	0.0988	0.0729	0.0526	0.0371
1.6	0.4731	0.5012	0.4825	0.4348	0.3725	0.3061	0.2426	0.1861	0.1385	0.1002	0.0705	0.0482	0.0320
1.8	0.7139	0.7018	0.6385	0.5481	0.4486	0.3524	0.2665	0.1945	0.1369	0.0929	0.0606	0.0377	0.0221
2.0	1.0347	0.9527	0.8205	0.6694	0.5207	0.3876	0.2762	0.1881	0.1217	0.0739	0.0412	0.0199	0.0069
2.2	1.4455	1.2523	1.0198	0.7866	0.5763	0.4004	0.2622	0.1597	0.0877	0.0400	0.0105	0.0060	0.0139
2.4	1.9477	1.5882	1.2175	0.8792	0.5958	0.3745	0.2122	0.1006	0.0293	0.0122	0.0329	0.0403	0.0399
-	-287	-263	-213	-154	-99	-52	-19	3	14	18	19	17	13

TABLE VI. (continued).

$\frac{x}{y}$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-2.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-2.2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-2.4	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

TABLE VII.—Real Part of  $\text{Bi}'(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x$ $y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.4483 246 0	0.4618 319 -10	0.5073 402 -21	0.5932 507 -39	0.7300 651 -61	0.9324 856 -92	1.2212 1155 -137	1.6266 1593 -204	2.1930 2237 -302	2.9855 3188 -454	4.1007 4600 -685	5.6815 6707 -1041	7.9418 9872 -1594
-0.2	0.4360 246 2	0.4459 315 -7	0.4873 391 -20	0.5680 488 -35	0.6977 620 -57	0.8900 811 -86	1.1641 1087 -127	1.5479 1492 -188	2.0824 2087 -279	2.8280 2964 -415	3.8735 4263 -624	5.3505 6195 -945	7.4549 9086 -1444
-0.4	0.3991 248 5	0.3986 302 -3	0.4283 361 -15	0.4943 434 -28	0.6039 534 -44	0.7673 680 -67	0.9993 1206 -99	1.3215 1663 -142	1.7654 2331 -209	2.3775 3106 -307	3.2253 3310 -454	4.4079 4751 -680	6.0713 6881 -1022
-0.6	0.3373 255 12	0.3210 287 2	0.3334 316 -7	0.3775 352 -16	0.4569 404 -27	0.5771 484 -39	0.7459 604 -55	0.9756 779 -78	1.2839 1033 -107	1.6964 1910 -150	2.2498 2212 -212	2.9959 2637 -298	4.0082 3667 -425
-0.8	0.2500 274 19	0.2148 274 9	0.2069 265 2	0.2256 255 -4	0.2698 248 -7	0.3387 248 -10	0.4327 258 -8	0.5525 276 -3	0.6999 297 6	0.8770 311 24	1.0850 300 58	1.3226 228 120	1.5819 33 228
-1.0	0.1350 311 24	0.0810 270 15	0.0539 215 8	0.0483 152 5	0.0579 83 10	0.0757 3 20	0.0936 -96 36	0.1018 -232 65	0.0862 -434 110	0.0262 -750 185	-0.1102 -1253 302	-0.3746 -2065 492	-0.8495 -3379 803
-1.2	-0.0112 372 27	-0.0799 280 15	-0.1208 174 9	-0.1443 56 12	-0.1624 -73 20	-0.1879 -223 40	-0.2361 -414 70	-0.3263 -676 116	-0.4851 -1057 189	-0.7511 -1630 302	-1.1827 -2510 476	-1.8693 -3875 752	-2.9497 -6005 1188
-1.4	-0.1949 460 23	-0.2690 304 9	-0.3128 140 3	-0.3426 -30 9	-0.3755 -209 24	-0.4295 -412 47	-0.5250 -664 88	-0.6877 -1006 145	-0.9522 -1494 230	-1.3679 -2215 357	-2.0081 -3299 554	-2.9838 -4947 856	-4.4593 -7463 1324
-1.6	-0.4247 571 12	-0.4887 337 -8	-0.5189 109 -11	-0.5380 -108 -5	-0.5680 -323 16	-0.6303 -553 45	-0.7482 -829 85	-0.9498 -1193 148	-1.2718 -1704 229	-1.7662 -2446 350	-2.5081 -3543 529	-3.6087 -5176 797	-5.2336 -7619 1206
-1.8	-0.7117 691 18	-0.7419 361 -39	-0.7357 66 -25	-0.7226 -190 -42	-0.7283 -421 -4	-0.7762 -650 29	-0.8893 -909 70	-1.0938 -1237 121	-1.4231 -1689 191	-1.9228 -2333 285	-2.6583 -3265 416	-3.7237 -4619 607	-5.2560 -6586 880
-2.0	-1.0678 793 24	-1.0310 345 -86	-0.9589 -17 -39	-0.8879 -300 -61	-0.8465 -525 -29	-0.8572 -720 4	-0.9400 -920 42	-1.1151 -1161 83	-1.4070 -1486 127	-1.8485 -1938 182	-2.4854 -2575 251	-3.3818 -4684 333	-4.6272 -6484 434
-2.2	-1.5025 821 154	-1.3538 241 -158	-1.1797 -182 -138	-1.0227 -470 -99	-0.9119 -658 -60	-0.8664 -785 -22	-0.8992 -889 10	-1.0209 -1004 37	-1.2434 -1155 58	-1.5819 -1362 68	-2.0571 -1636 64	-2.6965 -1974 42	-3.5336 -2348 42
-2.4	-2.0179 691 -281	-1.6994 -24 -258	-1.3811 -485 -204	-1.1097 -740 -145	-0.9111 -851 -89	-0.7968 -872 -44	-0.7695 -848 -13	-0.3267 -808 -2	-0.9648 -767 -6	-1.1796 -717 -35	-1.4658 -630 -97	-1.8142 -441 -213	-2.2050 -33 -412



TABLE VII. (continued).

$x$ $y$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.4483 246	0.4594 173	0.4877 94	0.5254 6	0.5635 -92	0.5924 -196	0.6017 -295	0.5817 -380	0.5239 -434	0.4232 -441	0.2788 -388	0.0962 -265	-0.1122 -72
-0.2	0.4360 246	0.4507 176	0.4830 99	0.5251 10	0.5681 -90	0.6022 -197	0.6165 -302	0.6008 -395	0.5458 -457	0.4455 -472	0.2985 -424	0.1098 -301	-0.1084 -102
-0.4	0.3991 248	0.4243 188	0.4683 116	0.5236 27	0.5816 -79	0.6316 -198	0.6617 -323	0.6596 -440	0.6137 -528	0.5155 -567	0.3612 -536	0.1540 -415	-0.0938 -199
-0.6	0.3373 255	0.3790 212	0.4418 150	0.5193 82	0.6028 -53	0.6808 -194	0.7393 -352	0.7626 -511	0.7349 -648	0.6429 -734	0.4781 -737	0.2406 -628	-0.0586 -388
-0.8	0.2500 274	0.3124 256	0.4002 209	0.5085 126	0.6291 -1	0.7493 -171	0.8520 -378	0.9168 -603	0.9214 -820	0.8444 -985	0.6697 -1052	0.3911 -974	0.0165 -711
-1.0	0.1350 311	0.2200 327	0.3373 306	0.4848 233	0.6551 97	0.8345 -110	1.0023 -383	1.1315 -704	1.1902 -1040	1.1452 -1333	0.9678 -1515	0.6405 -1508	0.1645 -1241
-1.2	-0.0112 372	0.0945 436	0.2435 454	0.4372 405	0.6708 266	0.9301 21	1.1905 -335	1.4166 -789	1.5633 -1298	1.5803 -1792	1.4190 -2189	1.0429 -2307	0.4391 -2084
-1.4	-0.1949 460	-0.0749 590	0.1036 670	0.3484 667	0.6590 544	1.0226 278	1.4114 -187	1.7800 -812	2.0663 -1568	2.1955 -2364	2.0891 -3080	1.6791 -3472	0.9260 -3395
-1.6	-0.4247 571	-0.3038 793	-0.1040 968	0.1919 1050	0.5915 978	1.0870 691	1.6495 140	2.2235 -697	2.7257 -1787	3.0479 -3028	3.0675 -4232	2.6665 -5128	1.7579 -5441
-1.8	-0.7117 691	-0.6122 1039	-0.4091 1359	-0.0709 1535	0.4244 1629	1.0801 1386	1.8712 761	2.7345 -315	3.5624 -1841	4.2033 -3715	4.4716 -5700	4.1716 -7417	3.1371 -8362
-2.0	-1.0678 793	-1.0248 1307	-0.8509 1839	-0.4937 2301	0.0921 2565	0.9313 2469	2.0131 1841	3.2732 533	4.5803 -1525	5.7291 -4265	6.4478 -7416	6.4256 -10467	5.3636 -12672
-2.2	-1.5025 821	-1.5678 1551	-1.4771 2383	-1.1482 3210	-0.4997 358	0.5312 4088	1.9651 3615	3.7523 2156	5.7450 -1244	7.6779 -4363	9.1664 -9190	9.7328 -14334	8.8719 -2151
-2.4	-2.0179 691	-2.2650 1683	-2.3417 2920	-2.1252 4286	-1.4808 5563	-0.2836 6410	1.5473 6387	4.0056 4999	6.9486 1805	10.0548 -3437	12.8010 -10675	14.4794 -18883	14.2712 -26925
	-281	-250	-140	80	421	865	1369	1824	2078	1946	1230	-207	-2354

TABLE VIII.—Imaginary part of  $\text{Be}''(x+iy)$ .Beneath each tabular entry are given  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  and  $\Delta_x^2 \Delta_y^2$ .

$x \backslash y$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.4
0.0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0	0.0000 0 0
0.2	0.0012 -70	-0.0270 -74	-0.0626 -89	-0.1073 -118	-0.1640 -166	-0.2374 -240	-0.3351 -350	-0.4683 -514	-0.6535 -757	-0.9155 -1120	-1.2909 -1665	-1.8352 -2489	-2.6318 -3745
0.4	0.0093 -133	-0.0466 -138	-0.1165 -164	-0.2030 -217	-0.3116 -306	-0.4512 -444	-0.6357 -649	-0.8858 -951	-1.2323 -1399	-1.7205 -2064	-2.4177 -3059	-3.4250 -4559	-4.8946 -6840
0.6	0.0307 -180	-0.0526 -181	-0.1541 -212	-0.2772 -282	-0.4289 -400	-0.6211 -581	-0.8722 -851	-1.2094 -1248	-1.6730 -214	-2.3218 -326	-3.2427 -498	-4.5653 -766	-6.4834 -1191
0.8	0.0698 -202	-0.0407 -193	-0.1709 -223	-0.3237 -299	-0.5069 -430	-0.7337 -633	-1.0246 -932	-1.4098 -1364	-1.9330 -283	-2.6577 -3907	-3.6766 -4252	-5.1260 -6246	-7.2082 -9227
1.0	0.1288 -192	-0.0099 -166	-0.1657 -188	-0.3407 -261	-0.5423 -392	-0.7837 -592	-1.0850 -881	-1.4753 -1291	-1.9662 -1874	-2.7065 -2707	-3.6905 -3911	-5.0702 -5662	-7.0228 -8231
1.2	0.2067 -138	0.0371 -92	-0.1421 -102	-0.3321 -167	-0.5392 -285	-0.7753 -463	-1.0583 -711	-1.4132 -1050	-1.8740 -1512	-2.4875 -2149	-3.3180 -3036	-4.4549 -4281	-6.0243 -6041
1.4	0.2979 -26	0.0929 39	-0.1088 37	-0.3072 -19	-0.5079 -122	-0.7211 -266	-0.9613 -453	-1.2471 -687	-1.6022 -979	-2.0557 -1346	-2.6447 -1811	-3.4159 -2396	-4.4378 -3126
1.6	0.3911 161	0.1443 235	-0.0797 232	-0.2809 175	-0.4648 83	-0.6406 -29	-0.8193 -149	-1.0129 -267	-1.2330 -372	-1.4901 -448	-1.7915 -464	-2.1384 -365	-2.5200 -49
1.8	0.4674 447	0.1716 507	-0.0742 481	-0.2723 404	-0.4302 307	-0.5673 217	-0.6625 154	-0.7519 142	-0.8264 208	-0.8792 395	-0.8908 774	-0.8226 1461	-0.6041 2680
2.0	0.4979 859	0.1473 867	-0.1173 779	-0.3044 651	-0.4263 525	-0.4955 434	-0.5207 408	-0.5044 474	-0.4396 671	-0.3062 1055	-0.0651 1717	-0.3512 2805	1.0535 4558
2.2	0.4412 1429	0.0357 1320	-0.2386 1116	-0.4014 894	-0.4746 705	-0.4766 588	-0.4190 570	-0.3033 680	-0.1183 952	0.1636 1444	0.5925 2248	1.2500 3508	2.2639 5456
2.4	0.2400 2185	-0.2087 1862	-0.4717 1469	-0.5876 1100	-0.5927 813	-0.5156 642	-0.3731 607	-0.1687 724	0.1096 1018	0.4915 1532	1.0291 2336	1.8035 3545	2.9373 5340
	199	66	-25	-86	-117	-135	-154	-174	-217	-288	-399	-578	-845

TABLE VIII. (continued).

$x \backslash y$	0.0	-0.2	-0.4	-0.6	-0.8	-1.0	-1.2	-1.4	-1.6	-1.8	-2.0	-2.2	-2.4
0.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.0012	0.0222	0.0369	0.0410	0.0369	0.0369	0.0222	-0.0022	-0.0365	-0.0781	-0.1232	-0.1664	-0.2008	-0.2193
-0.2	-70	-74	-83	-94	-103	-105	-97	-75	-36	19	86	160	228
0.0093	0.0519	0.0799	0.0914	0.0840	0.0840	0.0556	-0.0054	-0.0556	-0.1525	-0.2482	-0.3413	-0.4177	-0.4616
-0.4	-133	-144	-164	-190	-211	-220	-208	-166	-89	24	166	325	475
0.0307	0.0958	0.1404	0.1609	0.1524	0.1524	0.1109	0.0340	-0.0774	-0.2177	-0.3753	-0.5328	-0.6670	-0.7517
-0.6	-180	-203	-242	-289	-332	-357	-349	-293	-179	-2	229	495	759
0.0698	0.1598	0.2251	0.2592	0.2541	0.2541	0.2022	0.0979	-0.0596	-0.2644	-0.5017	-0.7469	-0.9659	-1.1181
-0.8	-202	-244	-311	-392	-470	-527	-538	-480	-333	-87	256	666	1093
0.1288	0.2480	0.3408	0.3967	0.3967	0.4030	0.3465	0.2163	0.0071	-0.2769	-0.6187	-0.9860	-1.3313	-1.5944
-1.0	-192	-261	-367	-496	-630	-742	-797	-759	-591	-269	209	816	1484
0.2067	0.3621	0.4929	0.5838	0.6152	0.6152	0.5655	0.4151	-0.131	-0.2291	-0.7074	-1.2449	-1.7777	-2.2193
-1.2	-138	-240	-396	-594	-811	-1013	-1151	-1108	-1005	-613	28	899	1915
0.2979	0.4997	0.6843	0.8301	0.8301	0.9085	0.8863	0.7301	0.4126	-0.0788	-0.7326	-1.5048	-2.3127	-3.0358
-1.4	-26	-165	-382	-670	-1005	-1345	-1625	-1759	-1632	-1216	-391	822	2338
0.3911	0.6532	0.9183	1.1430	1.3024	1.3024	1.3422	1.2087	0.8523	0.2394	0.6330	-1.7222	-2.9274	-4.0852
-1.6	161	-10	-295	-695	-1191	-1736	-2242	-2591	-2634	-2217	-1215	426	2634
0.4674	0.8068	1.1708	1.5245	1.8149	1.8149	1.9718	1.9128	1.5537	0.8248	-0.3068	-1.8127	-3.5798	-5.3951
-1.8	447	260	-88	-618	-1324	-2157	-3013	-3727	-4079	-3815	-2697	-567	2567
0.4979	0.9330	1.4356	1.9664	2.4588	2.4588	2.8169	2.9104	2.6308	1.8233	0.4082	-1.6251	-4.1671	-6.9552
-2.0	859	693	305	-359	-1320	-2543	-3915	-5220	-6142	-6282	-6218	-2605	1699
0.4412	0.9880	1.6676	2.4417	3.2324	3.2324	3.9150	4.3179	4.2331	3.4425	1.7614	-0.9025	-4.4795	-8.6726
-2.2	1429	1352	979	206	-1040	-2766	-4866	-7083	-8981	-9963	-9323	-6393	-729
0.2400	0.9054	1.7983	2.8923	4.1059	4.1059	5.2865	6.2019	6.5461	5.9674	4.1243	-0.7717	-4.1295	-10.2941
-2.4	2185	2305	2057	1256	-262	-2588	-5670	-9234	-1.2720	-1.5266	-1.5757	-1.2994	-5977
0.199	368	553	723	821	821	773	506	-50	-920	-2046	-3270	-4304	-4742

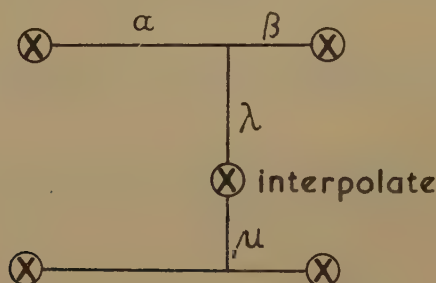
The tables may also be used for the region  $x = -2.4(0.2)2.4$ ,  $y = 0.0(0.2)2.4$  in the upper half of the complex plane (fig. 2) by taking complex conjugates.

### 3. Instructions for Interpolation.

The method of bivariate interpolation given in this section takes advantage of the fact that the real and imaginary parts of an analytic function of a complex variable each satisfy Laplace's differential equation. The resulting simplifications in the interpolation formula have governed the choice of differences which are tabulated. In order to perform interpolation by second and fourth differences in an ordinary function  $u$  of two real variables, it is normally necessary to tabulate the five central differences

$$\Delta_x^2 u, \Delta_y^2 u, \Delta_x^4 u, \Delta_y^4 u, \Delta_x^2 \Delta_y^2 u.$$

Fig. 3.



Co-ordinates of interpolate in the interpolation square.

In the present tables, it has been possible to arrange the interpolation formula so as to involve only the "coupled" differences

$$\frac{1}{2}(\Delta_x^2 - \Delta_y^2)u, \Delta_x^2 \Delta_y^2 u,$$

which, apart from constants of proportionality, approximate the second and fourth derivatives to a high degree of accuracy (see Appendix). These two quantities appear in the tables beneath each function value  $u$ .

The co-ordinates of the interpolate relative to the four nearest tabular entries are shown in fig. 3. The  $x$ -interval is divided in the ratio  $\alpha : \beta$ , and the  $y$ -interval in the ratio  $\lambda : \mu$ , such that

$$\alpha + \beta = \lambda + \mu = 1.$$

In performing an interpolation, each of the four function values at the corners of the interpolation square shown in fig. 3, together with the corresponding differences at these points, is multiplied by the coefficients shown in the following scheme. The products so formed are then added together to give the interpolated function value.



*Interpolation Coefficients.*

(a) Coefficients of function values :—

$\beta\mu$	$\alpha\mu$
$\beta\lambda$	$\alpha\lambda$

(b) Coefficients of  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  :—

$\frac{1}{6}\beta\mu(\beta^2 - \mu^2)$	$\frac{1}{6}\alpha\mu(\alpha^2 - \mu^2)$
$\frac{1}{6}\beta\lambda(\beta^2 - \lambda^2)$	$\frac{1}{6}\alpha\lambda(\alpha^2 - \lambda^2)$

(c) Coefficients of  $\Delta_x^2 \Delta_y^2$  :—

$-\frac{\beta\mu}{360}(4 + (\beta^2 - 3\mu^2)(3\beta^2 - \mu^2))$	$-\frac{\alpha\mu}{360}(4 + (\alpha^2 - 3\mu^2)(3\alpha^2 - \mu^2))$
$-\frac{\beta\lambda}{360}(4 + (\beta^2 - 3\lambda^2)(3\beta^2 - \lambda^2))$	$-\frac{\alpha\lambda}{360}(4 + (\alpha^2 - 3\lambda^2)(3\alpha^2 - \lambda^2))$

*Conditions for Neglect of Differences.*

Contribution from (c) : Negligible if modulus of each  $\Delta_x^2 \Delta_y^2$  is less than 40 in units of last decimal retained.

Contribution from (b) : Negligible if modulus of each  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  is less than 4 in units of last decimal retained.

*Simplified Second Difference Interpolation.*

If each pair of  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$  at diagonally opposite corners of the interpolation square differ (with due regard to sign) by less than 60, and if the modulus of each  $\Delta_x^2 \Delta_y^2$  is less than 40, the coefficients given at (b) may be replaced by the following :—

(b') Simplified coefficients of  $\frac{1}{2}(\Delta_x^2 - \Delta_y^2)$ —

$\frac{1}{4}\beta\mu(\beta - \mu)$	$\frac{1}{4}\alpha\mu(\alpha - \mu)$
$\frac{1}{4}\beta\lambda(\beta - \lambda)$	$\frac{1}{4}\alpha\lambda(\alpha - \lambda)$

These coefficients correspond to the use of Bessel's in place of Everett's formula for single variable interpolation. They cannot be used in conjunction with (c).

*Further Notes.*

Interpolation coefficients should be computed to one more decimal place than the number of significant figures in the corresponding differences.

The complete central difference interpolation formula includes all even differences, so that interpolation using (a), (b), and (c), in the above scheme presumes that sixth differences may be neglected. The contribution from sixth differences is negligible if they are less than 500 in units of the last decimal retained. For this purpose, any one of the quantities

$$\Delta_x^6 u, \Delta_x^4 \Delta_y^2 u, \Delta_x^2 \Delta_y^4 u, \Delta_y^6 u$$

may be reckoned as a sixth difference, since these are all approximately equal in magnitude by virtue of Laplace's equation.

4. *Accuracy of Tables.*

Even though every precaution as to accuracy were taken, any set of tables must have a rounding-off error up to a maximum of 0.5 in units of the last decimal. If tables are originally calculated to only one more decimal place than is finally required when rounded off, the rounding-off error is slightly greater. In the present instance, Tables I.-IV. were originally calculated to five decimal places with a probable error of about one in the fifth place. This would have given rise to a maximum error of something like 0.7 or 0.8 in units of the fourth place. A simple artifice, however, which will be described in the Appendix, was used in the process of rounding-off, by which it is hoped that this error has been reduced. Tables V.-VIII. were derived from Tables I.-IV. (before rounding-off), but here the loss of accuracy inevitable in such derivation has to some extent been reduced by making suitable use of the Cauchy-Riemann equations. To sum up, therefore, it is hoped that, apart from any typographical errors which may have passed undetected, the errors in Tables I.-IV. never exceed 0.6 or 0.7 and those of Tables V.-VIII. never exceed 0.8 in units of the fourth decimal.

5. *Appendix.*

In this section, the consequences of the Cauchy-Riemann equations are discussed with reference to interpolation, checking, and the calculation of the derived functions  $Ai'$  and  $Bi'$ .

Let

$$u + iv = f(x + iy). \quad (1)$$

The Cauchy-Riemann equations are

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}, \quad (2)$$

from which Laplace's equation is obtained :

$$\left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) (u \text{ or } v) = 0. \quad (3)$$



The coupled difference, in terms of horizontal differences only, is

$$\frac{1}{2}(\Delta_x^2 - \Delta_y^2) = \Delta_x^2 - \frac{1}{12} \Delta_x^4 + \dots,$$

whereas the "modified" horizontal second difference is

$$M_x^2 = \Delta_x^2 - 0.184 \Delta_x^4.$$

It is the modifying of the second difference by subtracting from it a suitable portion of the fourth difference which, by reducing the mean coefficient of the fourth difference, secures more rapid convergence. The modified second difference is a somewhat artificial quantity designed specifically for this purpose.

The fourth difference term in Pearson's formula involves the horizontal, vertical, and cross differences

$$\Delta_x^4 u, \Delta_y^4 u, \Delta_x^2 \Delta_y^2 u (= \Delta_y^2 \Delta_x^2 u),$$

Fig. 4.



Configuration of table-entries giving straight fourth differences at X.

which, to a first approximation, are equal to

$$D_x^4 u, D_y^4 u, D_x^2 D_y^2 u,$$

respectively. By Laplace's equation,

$$D_x^4 = D_y^4 = -D_x^2 D_y^2. \quad \dots \dots \dots (10)$$

It might seem, by analogy with the second difference, that the best choice for a coupled fourth difference would be

$$\frac{1}{2}(\Delta_x^4 + \Delta_y^4) = D^4 + \frac{1}{80} D^8 + \text{etc.}, \quad \dots \dots \dots (11)$$

where D without a subscript is used where the choice of subscript does not matter, by (3). The cross fourth difference, however, also has a symmetrical quality, and is more easily obtained. Its value depends on a more clustered set of function values (see figs. 4 and 5), which saves labour at the edges of the table and renders it more suitable for

Fig. 5.



Configuration of table-entries giving cross fourth difference at X.



interpolation purposes. For comparison with (11), it is given in terms of derivatives by

$$\Delta_x^2 \Delta_y^2 = -D^4 + \frac{1}{720} D^8 + \text{etc.}, \quad . . . . . (12)$$

and is therefore closer to the fourth derivative, in magnitude, than (11).

Re-arrangement of the interpolation formula in terms of the coupled second and cross fourth differences described above yields the coefficients set out in section 3.

### 5.2. Checking and Rounding-off Device for Analytic Functions.

A further result, which is both elegant and useful for checking purposes, may be derived from equations (6), (7), and (12). Adding (6) and (7), and making use of Laplace's equation,

$$\Delta_x^2 + \Delta_y^2 = \frac{1}{6} D^4 + \frac{1}{10,080} D^8 + \text{etc.}, \quad . . . . . (13)$$

which, by (12),

$$= -\frac{1}{6} \Delta_x^2 \Delta_y^2 + \frac{1}{3024} \Delta^8 + \text{etc.}, \quad . . . . . (14)$$

Hence, if terms of eighth and higher orders are negligible,

$$6(\Delta_x^2 + \Delta_y^2) + \Delta_x^2 \Delta_y^2 = 0. \quad . . . . . (15)$$

In terms of function values, this equation may be represented in the manner of section 3 by the following scheme :—

1	4	1	= 0. . . . .	(16)
4	-20	4		
1	4	1		

The numbers in the cells of the array are coefficients of a similar array of function values from the table, the products so formed being added together. It follows that

1/20	1/5	1/20	=	0	0	0
1/5	0	1/5		0	1	0
1/20	1/5	1/20		0	0	0

. . . . . (17)

which is an interpolation formula in Lagrangian form for a function value from the table, in terms of the values surrounding it on all sides. Its peculiar merit, as a checking device, lies in the fact that no differences less than the eighth are neglected in its derivation. A function value

interpolated in this way may indeed be more accurate than any of the eight values used to obtain it. For in a perfect table the rounding-off errors of the eight interpolants have uniform probability distributions between  $-0.5$  and  $0.5$  in units of the last decimal, and it may be shown that the probability of the resulting error of the interpolate is very nearly normally distributed with standard deviation  $0.12$ . This argument holds only if the spacing between function values is sufficiently large for the rounding-off errors to be virtually uncorrelated, yet sufficiently small for eighth differences to be negligible.

The above has been used as a check throughout the present tables, and has also been used to decide ambiguities in rounding off from five decimals to four.

### 5.3. Calculation of Derived Functions.

As before, let

$$f(x+iy)=u+iv.$$

Then

$$f'(x+iy)=\frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}=\frac{\partial v}{\partial y}-i\frac{\partial u}{\partial y}.$$

It is now necessary to introduce a notation for *central* first differences. An ordinary first difference is associated with a point half-way between two adjacent tabular positions, but the first difference at a tabular position, *i. e.* a central first difference, is defined as the average of the adjacent first differences on either side, or half the difference between the function values on either side. Such a difference will be denoted by  $\Delta$ . Then

$$\Delta=D+\frac{1}{6}D^3+\frac{1}{120}D^5+\frac{1}{5040}D^7+\frac{1}{362,880}D^9+\text{etc.}, \quad (18)$$

where the same subscript is understood throughout the equation. Hence, by the Cauchy-Riemann equations (2),

$$\left. \begin{aligned} \frac{1}{2}(\Delta_x u + \Delta_y v) &= D_x u + \frac{1}{120} D_x^5 u + \frac{1}{362,880} D_x^9 u + \dots \\ &= D_y v + \frac{1}{120} D_y^5 v + \frac{1}{362,880} D_y^9 v + \dots \end{aligned} \right\} \text{etc.} \quad (19)$$

This gives the real part of  $f'(x+iy)$  in terms of a coupled first difference together with fifth and higher order derivatives. The fifth derivative may be obtained in terms of the first difference of the cross fourth difference by applying (18) to (12) and using (2):—

$$\left. \begin{aligned} \frac{1}{2}(\Delta_x \Delta_x^2 \Delta_y^2 u + \Delta_y \Delta_x^2 \Delta_y^2 v) &= -D_x^5 u - \frac{1}{144} D_x^9 u - \dots \\ &= -D_y^5 v - \frac{1}{144} D_y^9 v - \dots \end{aligned} \right\} \dots \quad (20)$$

Hence

$$\left. \begin{aligned} D_x u &= D_y v = \frac{1}{2} (\Delta_x u + \Delta_y v) \\ &+ \frac{1}{240} (\Delta_x \Delta_x^2 \Delta_y^2 u + \Delta_y \Delta_x^2 \Delta_y^2 v) \\ &+ \frac{1}{18,144} \Delta^9, \text{ etc.} \\ &= h \cdot (\text{real part of } f'), \end{aligned} \right\}, \quad \cdot \cdot \cdot \cdot \quad (21)$$

where  $h$  is the argument interval defined previously. The imaginary part of  $f'$  is obtained in a similar way.

Use of the method described above to obtain tables of  $Ai'$  and  $Bi'$  from those of  $Ai$  and  $Bi$  introduces a degree of error depending on the choice of  $h$ , the tabular interval. It can be shown that if the probable error in the original five-decimal tables ( $h=0.2$ ) of  $Ai(z)$  is unity in the fifth place, the probable error in the derived table  $Ai'(z)$  is  $2\frac{1}{2}$  in the fifth place. This gives a probable maximum error of 0.75 in units of the fourth place in the rounded-off tables. The rounding-off artifice described in section 5.2 has probably reduced this maximum error considerably.

#### *Postscript.*

Since the preparation of these tables, there have been published\* eight-figure tables of the modified Hankel Functions of order one-third—functions which are closely related to the Airy functions. The tables cover the region

$$|x+iy| \leq 6$$

at intervals of 0.1 in  $x$  and  $y$ . No differences are tabulated, interpolation being performed either by Taylor's expansion or by the complex Lagrangian coefficients computed by Lowan and Salzer†. Interpolation by differences, however, possesses certain definite advantages in ease of manipulation, its chief being that the number of terms required to yield any given degree of accuracy may be seen at a glance. It was indeed by virtue of the popularity with computers of interpolation by Everett's and (especially) Bessel's formulæ that some extension of these methods for use in the complex plane has here been sought.

\* "Annals of the Computation Laboratory of Harvard University," volume ii., Harvard University Press, 1945.

† "Coefficients for Interpolation within a Square Grid in the Complex Plane," Jour. of Math. and Phys. xxiii, pp. 155-166 (1944).

XXVII. *The Determination of the Orientation of Single Crystals on a Weissenberg Goniometer.*

By NORA WOOSTER, M.A., Ph.D., and W. A. WOOSTER, M.A., Ph.D.,  
F.Inst.P., Brooklyn Crystallographic Laboratory, Cambridge \*.

[Received January 17, 1946.]

*Summary.*

A method is described by which a Weissenberg X-ray goniometer may be used to determine the orientation of a crystal whose cell size is known. The goniometer is used without the screens which normally cover most of the film, so that an unusual type of pattern is produced. The crystal is oscillated about the usual axis of the goniometer, and unfiltered radiation is employed, so that the photograph serves the same purpose as a combined oscillation and Laue photograph. Two charts are given for interpreting the patterns, and in Appendix I. the equations and values used in drawing up the charts are listed.

1. *Introduction.*

METHODS already described, by which the orientation of crystals may be found, include the use of the ionization spectrometer, Laue, and oscillation, photographs. Apart from the fact that ionization spectrometers are rare, the first method is limited by the relatively large size of crystal needed. The second method can be used successfully if the X-ray beam nearly coincides with an axis of crystallographic symmetry, but otherwise the labour of interpretation often becomes too great for the method to be practicable. The limitation inherent in oscillation photographs is that the result is liable to be indeterminate to the extent of the amplitude of oscillation.

By using a moving film, however, and the total radiation from a tube, one can eliminate the disadvantages and combine the advantages of both of these photographic methods. The characteristic radiation gives reflections which are easy to index. The white radiation gives rise to streaks, which pass through the characteristic spots, and these streaks can be interpreted to give the orientation of the crystal at the moment when the incident X-ray beam was parallel to the plane which gave rise to the reflection.

If  $\rho$  is the angle made by the normal to the reflecting plane with the axis of rotation, the white radiation streaks which are seen on the film

---

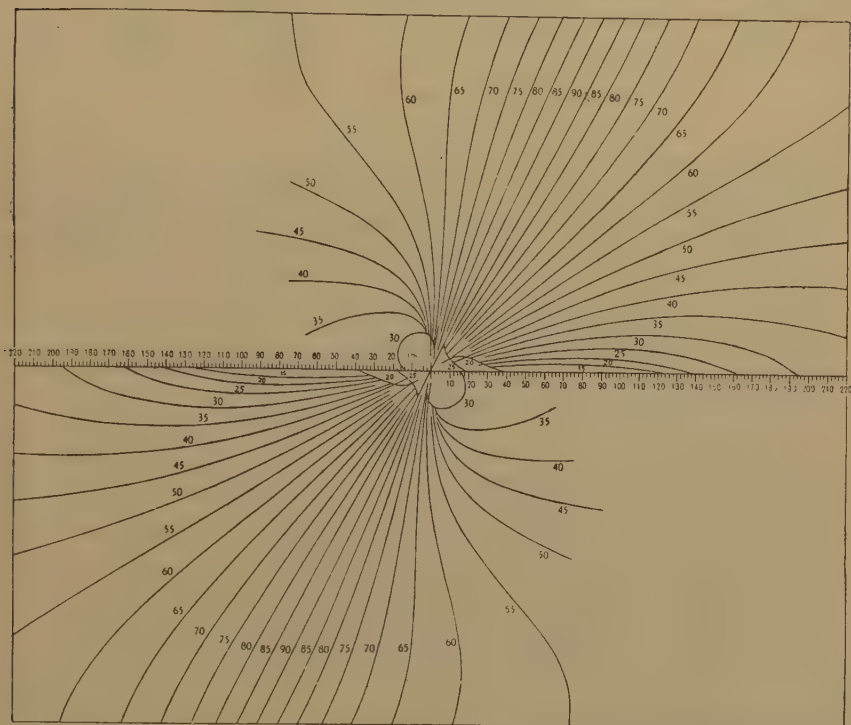
\* Communicated by the Authors.



are constant  $\rho$  curves \*. For each value of  $\rho$ , the curve has a characteristic shape. The curves for values of  $\rho$ , differing by  $5^\circ$ , plotted for a film moving 1 mm. for every  $2^\circ$  rotation, and of radius 3 cm., are drawn in Chart I.

The reflections for the characteristic radiation lie on constant  $\theta$  curves. On a stationary film these reflections lie on "powder lines," but on a moving film the corresponding lines are drawn out on one side, and compressed on the other. These curves, likewise at  $5^\circ$  intervals, are plotted on Chart II.

CHART I.†



In order to be able to make use of all the information which is imprinted on the film, it is necessary to define exactly the position of the crystal when the camera is at the extreme end of its traverse, and also the position of the film relative to the X-ray beam when the camera is in this position.

The position of the film in the camera is best recorded by means of two fiducial holes such that the line joining them is parallel to the edge of

---

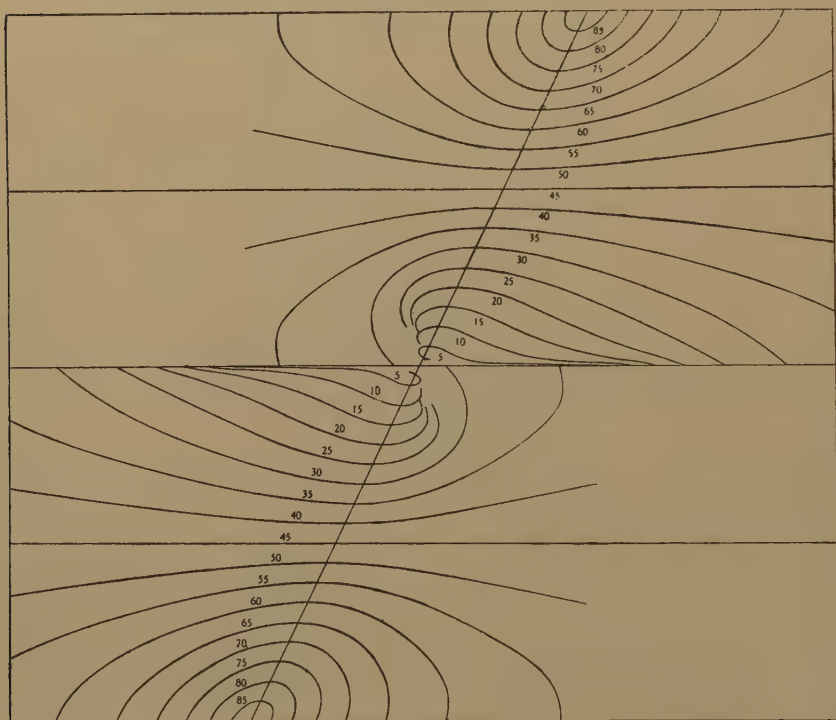
\* In the paper by J. D. Bernal, *Proc. Roy. Soc. A*, cxiii. p. 117 (1926),  $\rho$  is called  $\alpha$ . The nomenclature adopted in this paper is Buerger's. (See 'X-ray Crystallography,' by M. J. Buerger.)

† Full-sized printed copies of this chart may be obtained by arrangement with the authors.

the camera. When the exposure of the crystal is complete, the camera is removed and held so that the direct beam is allowed to fall on these holes, blackening the film underneath. These fiducial spots may then be placed on a line drawn on the charts at the appropriate distance from the centre, *i. e.* at a distance equal to that of the X-ray beam from this fiducial line.

The position of the crystal relative to the X-ray beam and the axis of the goniometer is defined in the following way. A certain face, either natural or artificial, is set parallel to the plane containing the axis of

CHART II.\*



the instrument, and the axis of the collimator of the X-ray beam. In addition, either the intersection of a second plane with the first must be set in a known direction, for instance, parallel to the axis of the goniometer, or a line drawn on this face may be set parallel to the same axis.

It is also necessary to define on the film the line in which it is cut by the plane perpendicular to the axis of the camera and passing through the incident X-ray beam. This line refers to the position of the camera carriage when it is at the extreme left-hand end of its traverse. When the camera is in this position a mark is made on it showing the inter-

---

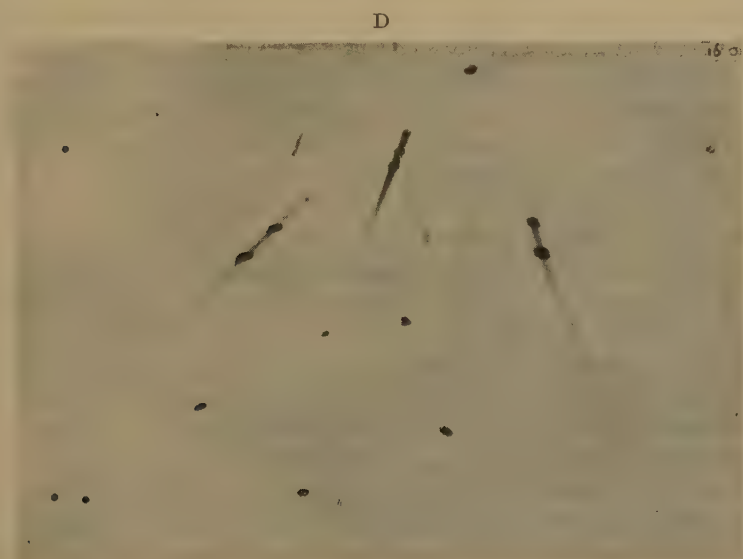
\* Full-sized printed copies of this chart may be obtained by arrangement with the authors.

section with it of a plane passing through the X-ray collimator and perpendicular to the axis of the instrument. The distance between this mark and one of the fiducial spots is measured, so that the mark, hereafter called D, may be easily transferred to the film.

## 2. Interpretation of the Photograph.

A typical photograph is reproduced in fig. 1. For interpretation it is superimposed on the charts, so that the upper fiducial spots lie on the line drawn parallel with the horizontal line through the centre of the charts, as described in paragraph 1. *Throughout all the operations it is essential to maintain this relative position.*

Fig. 1.



Unscreened Weissenberg photograph of a diamond in a diamond cutting tool. Angle of oscillation  $40^\circ$ . Full radiation from an X-ray tube with copper centre cathode.

In order to define the direction of the normal to the reflecting plane two polar co-ordinates are required,  $\rho$  and  $\omega$ .  $\rho$ , as mentioned above, is the angle made by the normal to the reflecting plane with the axis of rotation.  $\omega$  is the angle between two planes intersecting in the axis of rotation, one perpendicular to the direction of the incident X-rays and the other containing the normal to the reflecting plane. When the mark D on the film is set coincident with the vertical line through the point of convergence to the  $\rho$  curves on Chart I., the value of  $\omega$  is zero.

*Indexing the Reflecting Planes.*

The assignation of indices is a simple process. A table is drawn up containing the indices and  $\theta$  values of all the planes in the crystal under investigation which can reflect the characteristic radiation. The two charts may then be superposed and the curves for the  $\theta$  values which may occur are pricked through from Chart II. to Chart I. The points are then joined up by inked lines and the indices written in.

Then, in order to index a characteristic spot, the streak on which it occurs is placed on its proper  $\rho$  curve, *i. e.* the curve which has the same shape. The  $\theta$  value of the characteristic spot, and therefore its index, is immediately determined.

The straight curve on Chart I. is for  $\rho=90^\circ$ . On either side the curves are for diminishing values of  $\rho$ , and it is necessary for the interpretation to notice whether the relevant curve is to the right or to the left of the straight line. The value of  $\omega$  is given by the position of the mark D on the horizontal scale. Which set of values is taken to be positive and which negative is arbitrary, but it is necessary to be consistent. Throughout this paper the zero position of the camera is taken as the extreme left-hand position as judged by an observer looking in the direction in which the X-rays are travelling. As the camera travels to the other end of its traverse the crystal rotates, so that its upper surface moves away from the observer. The  $\omega$  values on the left of the centre of Chart I. are taken as positive, those on the right as negative.

On the chart all the  $\rho$  curves converge to one point, but on the photograph the streaks may be shifted into any parallel position. If the film needs to be shifted from its zero position to the left or to the right, to set any particular streak coincident with the corresponding  $\rho$  curve, then the amount the mark D moves is noted. This gives the value of  $\omega$ , which means that the crystal had to rotate from its initial position through that number of degrees before the normal to the reflecting plane was perpendicular to the X-ray beam. The reason for this may be understood if it is remembered that the shortest wave-lengths give rise to the part of the streak nearest the horizontal line. If the wave-length of the radiation became infinitely short the streak would be extrapolated back to the incident beam of X-rays, and the value of  $\theta$  would be  $0^\circ$ , *i. e.* the normal to the plane would be at right angles to the beam.

When a streak occurs without a characteristic reflection, it may still often, though not invariably, be indexed, for the index may sometimes be deduced. If the characteristic spots are absent it means that the crystal has not been oscillated into the reflecting position for the  $K\alpha$  and  $K\beta$  wave-lengths, in other words, that the streak has been cut off by the limited angle of oscillation. If the part of the streak which is present passes through the  $\theta$  curves for one or other of the planes which could reflect, then clearly the streak is not to be associated with any of these planes. Equally, there will be no streak where the characteristic spot would occur.

The accuracy with which the position of the streak may be determined



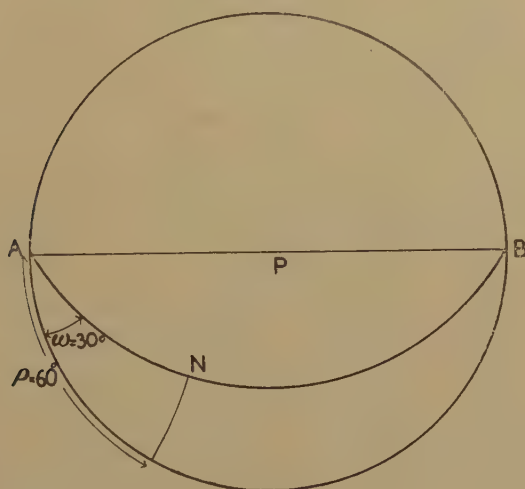
increases with its length, and is greater for large values of  $\rho$  than for small.

### 3. Transference of the Values of $\rho$ and $\omega$ for the Streaks to a Stereographic Projection.

The stereographic projection affords a convenient graphical method of determining the relative orientation of all the reflecting planes in the zero setting of the crystal. The  $\rho$  and  $\omega$  values of the streaks are listed, and if possible, indices assigned to them.

Suppose we consider a streak whose value for  $\rho$  is  $60^\circ$  L. and whose value for  $\omega$  is  $30^\circ$ . If the incident X-rays are supposed to fall vertically downwards on to the plane of the paper, then the corresponding normal to

Fig. 2.



Showing the method of transferring  $\rho$ ,  $\omega$  values to stereographic projection.

the reflecting plane is at N in the stereogram, fig. 2. Proceeding in this way all the normals to the planes which give rise to streaks in the photograph are plotted.

Since the angles between the normals to the planes and the crystallographic axes may be calculated, the position of the latter on the stereographic projection may be plotted. Small circles of appropriate radius are described about the normal to the reflecting planes, and the intersections of suitable circles fix the position of the axes.

#### *Example.*

This method has been extensively used for orientating diamonds, in particular, diamond cutting tools set in heavy steel shanks. A fitting was made so that the tool could be attached to the goniometer, since it was

much too heavy to fix to the ordinary crystal holder. The diamond cutting edge was on the axis of the instrument, and the length of the tool handle was at right angles to the same axis. The oscillation was limited by the size of the tool to  $40^\circ$ . The cylindrical camera with which the instrument (a two-crystal Weissenberg goniometer by Unicam Instruments, Ltd.) was fitted, was made in two pieces. Only the lower

TABLE I.

Reflecting plane	Bragg angle for			
	CuK $\beta$ 1.39 A.	CuK $\alpha$ 1.54 A.	FeK $\beta$ 1.74 A.	
111	$19\frac{1}{2}^\circ$	$22^\circ$	$25^\circ$	$28^\circ$
220	$33\frac{1}{2}$	$37\frac{1}{2}$	46	50
311	40	46	55	64
400	$50\frac{1}{2}$	60	80	
331	58	70		
422	$72\frac{1}{2}$			

TABLE II.

Indices of reflecting planes.	$\rho$ .	$\omega$ .
111	$85^\circ$ R	$18^\circ$
220	$63^\circ$ L	$34^\circ$
220	$58^\circ$ R	$42^\circ$
311	$90^\circ$	$50^\circ$
331	$67\frac{1}{3}^\circ$ R	$32^\circ$

$\rho$  and  $\omega$  values for the streaks on fig. 1.

TABLE III.

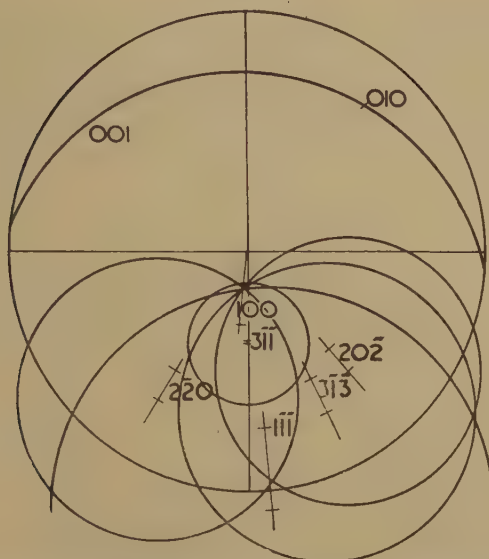
Indices of reflecting plane.	Angle on to cube axis.
111	$55^\circ$
220	$45^\circ$
311	$25^\circ$ , $72\frac{1}{2}^\circ$
331	$46\frac{1}{2}^\circ$ , $77^\circ$

Angles between normals to reflecting planes in fig. 1 and the crystallographic axes.

hemi-cylinder was used because of the obstruction offered by the tool shank.

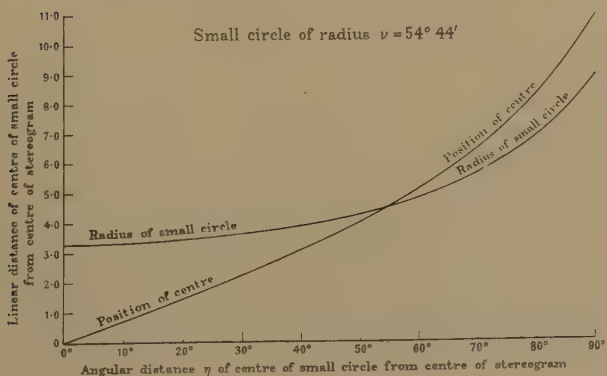
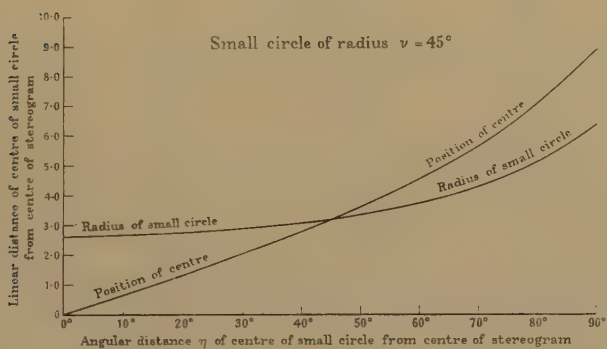
Table I., which contains all the planes in diamond which can reflect copper and iron radiation, was prepared and the curves for these  $\theta$  values

Fig. 3.



Stereograms of reflecting normals with indexed streaks and the corresponding directions of the cubic axes.

Fig. 4.



drawn on two  $\rho$  charts. The  $\rho$  and  $\omega$  values for the photograph given in fig. 1 are listed in Table II.

The angles between the planes which can reflect and the cube face are tabulated in Table III. Small circles are drawn round the normals to reflecting planes, and it will be seen in the stereogram reproduced in fig. 3, that four of the circles intersect within  $1^\circ$  in the point marked (100). The positions of (001) and (010) are deduced in the usual way. Generally

TABLE IV.

The distances, in centimetres, for the centre of the small circle from the centre of the projection, and the radius of the small circle, for values of  $\nu=54^\circ 44'$  ( $111 \wedge 100$ ) and  $\nu=45^\circ$  ( $110 \wedge 100$ ) for any cubic crystal.

$\eta$ .	$\nu=54^\circ 44'$ .		$\nu=45^\circ$ .	
	Distance of centre of small circle from the centre of the projection.	Radius of small circle.	Distance of centre of small circle from the centre of the projection.	Radius of small circle.
	cm.	cm.	cm.	cm.
$0^\circ$	0.00	3.29	0.00	2.63
$10^\circ$	0.71	3.32	0.65	2.66
$20^\circ$	1.44	3.42	1.31	2.72
$30^\circ$	2.21	3.60	2.02	2.86
$40^\circ$	3.05	3.87	2.77	3.05
$50^\circ$	3.99	4.25	3.61	3.33
$60^\circ$	5.11	4.82	4.56	3.72
$70^\circ$	6.49	5.64	5.65	4.29
$80^\circ$	8.32	6.90	7.11	5.11
$90^\circ$	11.01	8.98	8.99	6.36

speaking two octahedral planes are sufficient to fix the axes, or a combination of one octahedral, one dodecahedral and one (311) or (331) plane. Two dodecahedral planes are not enough.

It saves a considerable amount of time to prepare curves giving the position of the centre and the radius of small circles for drawing about (111) and (220) planes (see fig. 4). Values for these curves are given in Table IV. and the method of calculation is given in Appendix II.

Finally, the positions in the stereogram of the two reference faces mentioned in paragraph 1 must also be marked in. The usual solution to the problem of orientating the crystal is expressed in terms of the angle between these two faces and the cube axes; or it may be given in terms of the angles between the reference faces and the normals to the octahedral or dodecahedral planes.

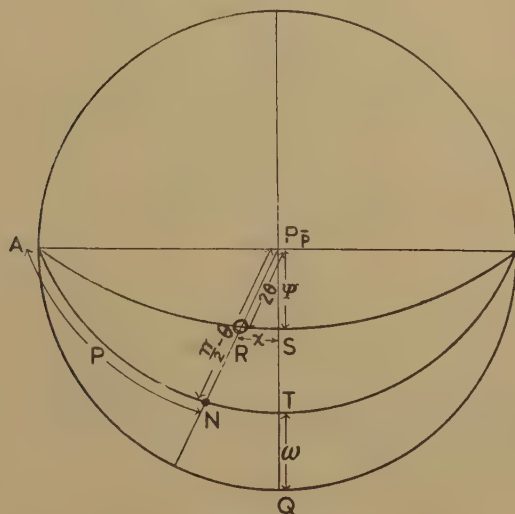


## APPENDIX I.

 Calculation of the  $\theta$  and  $\rho$  Curves for an unscreened Weissenberg Photograph.

In fig. 5 the angles between the directions which have to be considered

Fig. 5.



Stereogram showing the Naperian triangles on which the formulae for calculating  $\rho$  and  $\theta$  curves are based.

are assigned certain symbols, based on the conventions adopted in 'X-Ray Crystallography,' by M. J. Buerger.

P represents the direction of the incident X-rays, considered as travelling vertically downwards into the paper.

A represents the axis of rotation.

N represents the normal to the reflecting plane.

R represents the direction of the reflected beam.

Q represents a direction perpendicular to A and P.

As shown in the stereogram fig. 5,

$$\widehat{AN} = \rho,$$

$$\widehat{QAN} = \omega,$$

$$\widehat{NP} = \pi/2 - \theta,$$

$$\widehat{RP} = 2\theta \text{ (where } \bar{P} \text{ is diametrically opposite P),}$$

$$\widehat{RAP} = \gamma,$$

$$\widehat{RA} = \pi/2 - \chi.$$

A number of relations between these quantities follow from the solution of the Napierian triangles :

In triangle  $RSP$ , we have

$$\cos 2\theta = \cos \chi \cos Y. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\cos \overset{\Delta}{SPR} = \tan Y \cot 2\theta. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\sin Y = \tan \chi \cot \overset{\Delta}{SPR}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\sin \chi = \sin \overset{\Delta}{SPR} \sin 2\theta. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

In triangle  $NTP$  :

$$\sin \theta = \sin \rho \sin \omega. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\cos \overset{\Delta}{NPT} = \tan \theta \cot \omega. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\cos \omega = \cot \rho \cot \overset{\Delta}{TPN}. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\cos \rho = \sin \overset{\Delta}{TPN} \cos \theta. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

As may be seen from the stereogram  $\overset{\Delta}{TPN} = \overset{\Delta}{SPR}$  and combining equations (2) and (6), we have

$$\begin{aligned} \tan Y \cot 2\theta &= \tan \theta \cot \omega \\ \text{or } \tan \omega &= \frac{\tan \theta \tan 2\theta}{\tan Y} \quad . \quad . \quad . \quad . \quad . \quad (9) \end{aligned}$$

#### *Relation between Linear Co-ordinates ( $x, z$ ) on the Film and Angular Co-ordinates ( $\rho, \omega$ ).*

The linear co-ordinate  $z$  is measured along the film from an arbitrary zero in the direction of the axis of rotation. The linear co-ordinate  $x$  is measured from the line in which the direct beam strikes the film in a direction perpendicular to the axis of rotation. If  $r$  is the radius of the cylindrical camera

$$z = r \tan \chi$$

when  $z$  is measured from the point where the X-rays strike the film at the instant of reflection.

Also,

$$x = rY.$$

If  $z$  is measured from the point where the X-rays strike the film at the position for which  $\omega = 0$ , then to the above value of  $z$  must be added or subtracted a quantity  $k\omega$  where  $k$  is the translation of the camera for unit angular rotation of the crystal. Thus, as used in this paper,

$$z = r \tan \chi \pm k\omega. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$x = rY. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

### Constant $\rho$ Curves.

Combining equations (4), (5), (8), we have

$$\sin \chi = \frac{\cos \rho}{\cos \theta} \cdot 2 \sin \rho \sin \omega \cos \theta = \sin 2\rho \sin \omega. \quad (12)$$

Combining equations (3) and (7) we have

$$\frac{\cos \omega}{\cot \rho} = \frac{\sin Y}{\tan \chi}$$

$$\sin Y = \cos \omega \tan \rho \tan \chi. \quad (13)$$

If given values are assigned to  $\rho$  and  $\omega$ , then  $\chi$  may be found from equation (12) and  $Y$  from equation (13). Inserting these values of  $\chi$  and  $Y$  in equations (10) and (11) we obtain  $z$  and  $x$ . In this way the table of values for  $z$  and  $x$  for given values of  $\rho$ , Table V., has been calculated, and the results are plotted for a camera of radius 3.00 cm. in Chart I.

### Constant $\theta$ curves.

To trace out the curves for given values of  $\theta$ , a particular value of  $\theta$  is combined with an arbitrary value of  $\omega$ . From equation (5) we obtain the corresponding value of  $\rho$ ; from equation (12) the value of  $\chi$  and from equation (9) the value of  $\Delta$ . Using equations (10) and (11) we obtain the corresponding values of  $x$  and  $z$ . These are given in Table VI. and are plotted in Chart II.

## APPENDIX II.

### *The Calculation of the Radius and the Position of the Centre of a Small Circle of given Radius described about any Point in the Stereogram.*

If a small circle of radius  $\nu^\circ$  is to be described about a point W,  $\eta^\circ$  from the centre of a stereogram of radius R, we know that the linear distance of the ends of the diameter of the small circle passing through W from the centre of the stereogram are  $R \tan \frac{(\nu+\eta)}{2}$  and  $R \tan \frac{(\nu-\eta)}{2}$  respectively.

If S is the linear radius of the small circle we have

$$S = \frac{1}{2}R \left( \tan \frac{(\nu+\eta)}{2} - \tan \frac{(\nu-\eta)}{2} \right).$$

If T is the linear distance of the centre of the small circle from the centre of the stereogram,

$$T = \frac{1}{2}R \left( \tan \frac{(\nu+\eta)}{2} + \tan \frac{(\nu-\eta)}{2} \right).$$

Table IV. gives corresponding values of  $\eta$ , T and S calculated according to the above formulas, where R is  $2\frac{1}{2}$ ".

The figures in Tables V. and VI. refer to the upper halves of the Charts I. and II. Columns marked "Left" and "Right" refer to points on the left and the right respectively of the diagonal line.

TABLE V.  
 $\rho$  Curves for Moving Film.

5°			10°			15°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
3.97	0	5.03	3.41	0	5.59	2.77	0	6.23
3.72	0	4.78	3.16	.02	5.34	2.53	.04	5.97
3.48	.05	4.52	2.93	.03	5.07	2.30	.08	5.70
			2.70	.05	4.80	2.09	.12	5.41
			2.48	.06	4.52	1.90	.15	5.10
			2.27	.07	4.23	1.72	.17	4.78
			2.07	.08	3.93	1.56	.19	4.44
						1.40	.21	4.10
						1.26	.22	3.74
						1.12	.22	3.38
						.98	.21	3.02
						.85	.20	2.65
						.73	.18	2.27
						.60	.16	1.90
						.48	.13	1.52
20°			25°			30°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
1.98	0	7.02	.92	0	8.08	— .70	0	9.70
1.74	.08	6.76	.71	.14	7.79	— .87	.26	9.37
1.55	.16	6.45	.54	.28	7.46	— .90	.51	8.90
1.37	.22	6.13	.44	.40	7.06	— .83	.69	8.33
1.22	.28	5.78	.39	.50	6.61	— .71	.84	7.71
1.10	.33	5.40	.36	.57	6.14	— .55	.94	7.05
.99	.37	5.01	.34	.63	5.66	— .40	1.00	6.40
.89	.39	4.61	.33	.65	5.17	— .27	1.02	5.77
.80	.40	4.20	.32	.66	4.68	— .16	1.00	5.16
.72	.40	3.78	.32	.64	4.18	— .07	.97	4.57
.64	.38	3.36	.30	.61	3.70	— .01	.90	4.01
.56	.36	2.94	.28	.56	3.22	.03	.82	3.47
.48	.32	2.52	.26	.50	2.74	.06	.73	2.94
.40	.28	2.10	.22	.45	2.28	.07	.62	2.43
.32	.23	1.68	.19	.36	1.81	.07	.51	1.93
.24	.18	1.26	.14	.26	1.36	.06	.39	1.44
			.10	.19	.90	.04	.28	.96



TABLE V. (cont.).

35°			40°			45°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
—3.34	.90	11.34	—3.83	2.33	10.83	—4.74	3.66	11.74
—2.73	1.21	10.23	—2.69	2.35	9.19	—3.19	3.40	9.69
—2.14	1.40	9.14	—1.90	2.26	7.90	—2.20	3.14	8.20
—1.63	1.50	8.13	—1.34	2.15	6.84	—1.54	2.88	7.04
—1.20	1.54	7.20	— .94	2.01	5.94	—1.08	2.62	6.08
— .86	1.52	6.36	— .66	1.84	5.16	— .75	2.36	5.25
— .61	1.46	5.61	— .46	1.66	4.46	— .52	2.10	4.52
— .42	1.37	4.92	— .30	1.47	3.80	— .35	1.83	3.85
— .27	1.26	4.27	— .20	1.27	3.20	— .23	1.57	3.23
— .17	1.13	3.67	— .12	1.07	2.62	— .15	1.31	2.65
— .10	.99	3.10	— .07	.86	2.07	— .09	1.05	2.09
— .05	.83	2.55	— .04	.65	1.54	— .05	.79	1.55
— .02	.68	2.02	— .02	.43	1.02	.03	.52	1.03
0	.51	1.50	.01	.20	.49			
0	.34	1.00						

50°			55°			60°		
Left 3 tan $\chi$ cm.	3Y c.m.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
—3.83	4.99	10.83	—3.75	9.43	12.75	— .70	9.43	9.70
—2.69	4.45	9.19	—3.73	8.41	12.23	— .87	8.66	9.37
—1.90	4.02	7.90	—3.34	7.47	11.34	— .90	7.88	8.90
—1.34	3.61	6.84	—2.73	6.64	10.23	— .83	7.16	8.33
— .94	3.22	5.94	—2.14	5.93	9.14	— .71	6.49	7.71
— .66	2.87	5.16	—1.63	5.30	8.13	— .55	5.86	7.05
— .46	2.53	4.46	—1.20	4.74	7.20	— .40	5.28	6.40
— .30	2.19	3.80	— .86	4.24	6.36	— .27	4.73	5.77
— .20	1.87	3.20	— .61	3.78	5.61	— .16	4.23	5.16
— .12	1.55	2.62	— .42	3.35	4.92	— .07	3.75	4.57
— .07	1.24	2.07	— .27	2.93	4.27	— .01	3.29	4.01
— .04	.93	1.54	— .17	2.54	3.67	.03	2.84	3.47
— .02	.61	1.02	— .10	2.16	3.10	.06	2.41	2.94
			— .05	1.78	2.55	.07	1.98	2.43
			— .02	1.42	2.02	.07	1.59	1.93
			0	1.06	1.50	.06	1.19	1.44
			0	.70	1.00	.04	.79	.96
			0	.35	.50			



TABLE VI.  
 $\theta$  Curves for Moving Film.

5°			10°			15°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
.25	.52	.25	.50	1.05	.50	.75	1.57	.75
.02	.45	.55	.25	1.02	.78	.50	1.55	1.02
— .01	.31	.84	.05	.91	1.10	.27	1.49	1.33
.09	.23	1.04	— .05	.70	1.56	.07	1.38	1.67
.28	.16	1.29	.02	.57	1.86	— .09	1.19	2.10
	0	5.03		.34	2.53	— .01	.90	2.73
				0	5.59		.71	3.16
							.42	4.05
							.30	4.55
							0	6.23

20°			25°			30°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
1.00	2.09	1.00	1.25	2.62	1.25	1.50	3.14	1.50
.75	2.08	1.27	.99	2.61	1.52	1.24	3.14	1.77
.51	2.04	1.56	.75	2.58	1.81	1.00	3.12	2.05
.28	1.97	1.89	.51	2.53	2.12	.76	3.08	2.36
.07	1.85	2.26	.28	2.45	2.46	.52	3.03	2.70
— .12	1.69	2.68	.06	2.35	2.86	.28	2.96	3.07
— .23	1.45	3.24	— .15	2.20	3.31	.03	2.87	3.50
— .15	1.08	4.05	— .34	2.01	3.86	— .22	2.74	3.98
	.86	4.56	— .49	1.72	4.55	— .48	2.58	4.55
	.51	5.46	— .47	1.29	5.53	— .75	2.36	5.25
	.36	5.89	— .34	1.02	6.09	— 1.02	2.04	6.13
	0	7.02		.60	6.95	— 1.25	1.54	7.32
				.42	7.30	— 1.28	1.22	7.96
				0	8.08	— 1.20	.72	8.79
						— 1.11	.52	9.07
						— .70	0	9.70

TABLE VI. (cont.).

35°			40°			45°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
1.75	3.67	1.75	2.00	4.19	2.00	2.30	4.71	2.30
1.49	3.66	2.02	1.74	4.19	2.27	1.80	"	3.00
1.24	3.65	2.30	1.49	4.18	2.55	1.30	"	3.70
1.00	3.63	2.61	1.25	4.17	2.85	.80	"	4.40
.75	3.59	2.94	1.00	4.16	3.18	.30	"	5.10
.50	3.55	3.30	.75	4.14	3.54	—20	"	5.80
.25	3.49	3.71	.48	4.11	3.94	—70	"	6.50
—02	3.42	4.18	.19	4.07	4.40	—1.20	"	7.20
—33	3.32	4.71	—12	4.03	4.91	—1.70	"	7.90
—66	3.20	5.34	—48	3.97	5.52	—2.20	"	8.60
—1.06	3.03	6.10	—97	3.89	6.18	—2.70	"	9.30
—1.53	2.80	7.04	—1.46	3.79	7.11	—3.20	"	10.00
—2.15	2.45	8.25	—2.18	3.65	8.22	—3.70	"	10.70
—2.96	1.88	9.91	—3.18	3.44	9.69	—4.20	"	11.40
—3.37	1.51	10.77	—4.73	3.12	11.76	—4.70	"	12.10
—3.78	.91	11.85						
—3.82	.64	12.23						
—3.74	0	12.75						

50°			55°			60°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
2.50	5.23	2.50	2.75	5.76	2.75	3.00	6.28	3.00
2.24	5.23	2.77	2.49	5.76	3.02	2.74	6.29	3.27
1.99	5.24	3.05	2.25	5.77	3.30	2.50	6.31	3.56
1.75	5.25	3.35	2.00	5.79	3.61	2.25	6.34	3.86
1.50	5.27	3.68	1.75	5.83	3.94	2.01	6.39	4.20
1.24	5.29	4.04	1.50	5.87	4.30	1.76	6.47	4.56
.98	5.32	4.44	1.24	5.93	4.70	1.50	6.56	4.97
.68	5.35	4.88	.95	6.00	5.15	1.23	6.68	5.43
.36	5.40	5.40	.63	6.10	5.67	.93	6.84	5.96
—01	5.46	5.99	.26	6.22	6.26	.58	7.07	6.58
—50	5.53	6.65	—18	6.39	6.97	.18	7.38	7.34
—1.03	5.63	7.54	—73	6.63	7.84	—32	7.89	8.25
—1.79	5.77	8.61	—1.46	6.97	8.94	—53	8.20	8.71
—2.84	5.98	10.03	—2.46	7.54	10.41	—73	8.70	9.26
—4.45	6.31	12.04	—2.97	7.91	11.17	—77	8.90	9.42
			—3.54	8.51	12.09	—70	9.42	9.70
			—3.67	8.78	12.38			
			—3.74	9.42	12.75			



TABLE VI. (cont.).

65°			70°			75°		
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.
3.25	6.81	3.25	3.50	7.33	3.50	3.75	7.85	3.75
3.00	6.82	3.52	3.25	7.34	3.77	3.50	7.87	4.03
2.74	6.84	3.81	3.00	7.38	4.07	3.26	7.93	4.33
2.51	6.90	4.12	2.78	7.46	4.38	3.05	8.05	4.66
2.28	6.97	4.46	2.55	7.57	4.74	2.86	8.23	5.04
2.04	7.08	4.84	2.34	7.73	5.13	2.70	8.52	5.49
1.79	7.22	5.25	2.13	7.97	5.59	2.65	8.71	5.71
1.54	7.42	5.74	1.94	8.34	6.14	2.65	9.01	5.98
1.26	7.70	6.30	1.88	8.57	6.40	2.67	9.13	6.06
1.00	8.14	6.99	1.86	8.92	6.72	2.77	9.42	6.23
.89	8.40	7.32	1.88	9.07	6.82			
.83	8.82	7.73	1.98	9.42	7.02			
.83	9.00	7.86						
.92	9.42	8.08						

80°			85°					
Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.	Left 3 tan $\chi$ cm.	3Y cm.	Right 3 tan $\chi$ cm.			
4.00	8.38	4.00	4.25	8.90	4.25			
3.76	8.41	4.28	4.02	8.97	4.55			
3.54	8.51	4.61	3.93	9.11	4.77			
3.37	8.72	4.98	3.97	9.42	5.03			
3.33	8.86	5.16						
3.32	9.08	5.38						
3.41	9.42	5.59						

# XXVIII. *Theory of Diffraction.*

By F. B. PIDDUCK\*.

[Received August 29, 1945.]

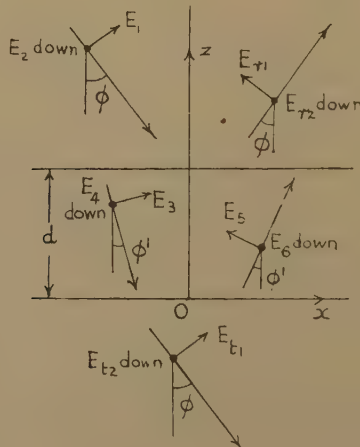
## 1. *Introduction.*

THIS paper contains a fuller account of a theory given more briefly in a book on 'Currents in Aerials and High-Frequency Networks' to be published in 1946 by the Clarendon Press, Oxford. The expression (14) below is Kirchhoff's integral in the case in which the source, obstacle, and observer are nearly in the same straight line. Kirchhoff's theory, like those of Fresnel and Huygens which preceded it, assumes that a light wave can act in a direction different from that in which it is travelling. But that can only happen if it excites a Hertzian oscillator which can radiate in all directions. It seems, therefore, that some considerations like the present must ultimately replace those theories.

## 2. *Electric Wave falling on a Dielectric Sheet.*

Let the wave be plane and the sheet an infinite sheet between the planes  $z=0$  and  $z=d$  in fig. 1. We use the m.k.s. system of electrical

Fig. 1.



units adopted by the International Electrotechnical Commission in 1938. The incident electric wave has components

$$(E_1 \cos \phi, E_2, E_1 \sin \phi) \exp (i\omega t - ikx \sin \phi + ikz \cos \phi), \quad \dots (1)$$

\*Communicated by the Author.

where  $\phi$  is the angle of incidence and the real part of the complex expressions is taken in the physical interpretation. Only the amplitudes are shown in fig. 1. The waves are polarized in the plane of incidence when  $E_2=0$ , and at right angles to it when  $E_1=0$ . The constant  $k$  is  $2\pi/\lambda$ , where  $\lambda$  is the wave-length in free space. The sheet can be taken to have a complex dielectric constant  $K=\epsilon-i\gamma/\omega\epsilon_0$ , where  $\gamma$  is the conductivity and  $\epsilon_0=8.854\times 10^{-12}$ . Thus

$$K=\epsilon-i\eta, \text{ where } \eta=1.129\times 10^{11}\gamma/\omega. \quad (2)$$

The wave refracted downwards in the sheet has components

$$(E_3 \cos \phi', E_4, E_3 \sin \phi') \exp(i\omega t - ik'x \sin \phi' + ik'z \cos \phi'),$$

where

$$k'=k(\epsilon-i\eta)^{\frac{1}{2}} \quad (3)$$

and  $\phi'$  is the complex angle of refraction (shown as real in fig. 1) given by

$$k' \sin \phi' = k \sin \phi. \quad (4)$$

The square root in (3), and that implied in  $\cos \phi'$ , are to be taken so that their real parts are positive.

The rest of the notation is shown in fig. 1, in which the waves are the *complete* waves obtained by adding the results of multiple reflection and refraction. The amplitudes of the reflected wave are  $E_{r1}$ ,  $E_{r2}$  corresponding to  $E_1$ ,  $E_2$ , those of the transmitted wave  $E_{t1}$ ,  $E_{t2}$ , and of the two waves in the sheet  $E_3$ ,  $E_4$  and  $E_5$ ,  $E_6$ . Their calculation is given in books on physical optics and involves no point of novelty, the results being transferable immediately from  $K$  real to  $K$  complex. There are eight boundary conditions expressing the continuity of the tangential components of electric and magnetic intensity at the two interfaces. If we write

$$\left. \begin{aligned} \Delta_1 &= (k' \cos \phi + k \cos \phi')^2 \exp(ik'd \cos \phi') \\ &\quad - (k' \cos \phi - k \cos \phi')^2 \exp(-ik'd \cos \phi'), \\ \Delta_2 &= (k' \cos \phi' + k \cos \phi)^2 \exp(ik'd \cos \phi') \\ &\quad - (k' \cos \phi' - k \cos \phi)^2 \exp(-ik'd \cos \phi'), \end{aligned} \right\} \quad (5)$$

the reflected and transmitted intensities are given by

$$\left. \begin{aligned} \Delta_1 E_{r1}/E_1 &= (k'^2 \cos^2 \phi - k^2 \cos^2 \phi') [\exp(ik'd \cos \phi') \\ &\quad - \exp(-ik'd \cos \phi')] \exp(2ikd \cos \phi), \\ \Delta_2 E_{r2}/E_2 &= -(k'^2 - k^2) [\exp(ik'd \cos \phi') \\ &\quad - \exp(-ik'd \cos \phi')] \exp(2ikd \cos \phi), \\ \Delta_1 E_{t1}/E_1 &= \Delta_2 E_{t2}/E_2 = 4kk' \cos \phi \cos \phi' \exp(ikd \cos \phi), \end{aligned} \right\} \quad (6)$$

and the fields in the sheet by

$$\left. \begin{aligned} \Delta_1 E_3/E_1 &= 2k \cos \phi (k' \cos \phi + k \cos \phi') \exp(ikd \cos \phi), \\ \Delta_1 E_5/E_1 &= -2k \cos \phi (k' \cos \phi - k \cos \phi') \exp(ikd \cos \phi), \\ \Delta_2 E_4/E_2 &= 2k \cos \phi (k' \cos \phi' + k \cos \phi) \exp(ikd \cos \phi), \\ \Delta_2 E_6/E_2 &= 2k \cos \phi (k' \cos \phi' - k \cos \phi) \exp(ikd \cos \phi). \end{aligned} \right\} \quad (7)$$

### 3. Interpretation in Terms of Electric Moment.

We can show that the field above and below the sheet is such as would be produced by the electric moments in the sheet. This step is not necessary, but desirable in view of what follows. The electric moment per unit volume is  $\epsilon_0(K-1)$  times the electric intensity, or at the point  $(\xi, \eta, \zeta)$  in the sheet

$$\left. \begin{aligned} P_\xi &= \epsilon_0(k'^2/k^2 - 1) \cos \phi' [E_3 \exp(i\omega t - ik\xi \sin \phi + ik'\zeta \cos \phi') \\ &\quad - E_5 \exp(i\omega t - ik\xi \sin \phi - ik'\zeta \cos \phi')], \\ P_\eta &= \epsilon_0(k'^2/k^2 - 1) [E_4 \exp(i\omega t - ik\xi \sin \phi + ik'\zeta \cos \phi') \\ &\quad + E_6 \exp(i\omega t - ik\xi \sin \phi - ik'\zeta \cos \phi')], \\ P_\zeta &= \epsilon_0(k'^2/k^2 - 1) \sin \phi' [E_3 \exp(i\omega t - ik\xi \sin \phi + ik'\zeta \cos \phi') \\ &\quad + E_5 \exp(i\omega t - ik\xi \sin \phi - ik'\zeta \cos \phi')]. \end{aligned} \right\} \quad (8)$$

From the polarization we get Hertz's vector and the electric intensity by the formulæ

$$\Pi = \int [P] dv / 4\pi\epsilon_0 r, \quad \mathbf{E} = \nabla(\nabla \cdot \Pi) + k^2 \Pi. \quad (9)$$

Verification requires the integration formula

$$\begin{aligned} &\int \exp(-ik\xi \sin \phi - ikr) dS/r \\ &= 2\pi i \exp(-ikx \sin \phi - ikz \cos \phi - ik|z - \zeta| \cos \phi) / k \cos \phi, \end{aligned}$$

where  $r$  is the distance from the point  $(x, y, z)$  to  $(\xi, \eta, \zeta)$ ,  $dS = d\xi d\eta$  is an element of area of a plane  $\zeta = \text{const.}$ , and integration is over the whole of that plane. This is proved by putting  $\xi = x + \rho \cos \theta$ ,  $\eta = y + \rho \sin \theta$  and using a modification of one of Sonine's integrals in Bessel's functions.

The reflected electric intensity, and the difference between the transmitted and incident electric intensities, are due to the polarization of the sheet. If we realize this, and particularly that the sheet may act obliquely, we shall be ready for the next step.

### 4. Diffraction by a Small Dielectric Disk.

In what follows the wave-length  $\lambda$  of the incident radiation is a small fraction of the linear dimensions of the disk, though not necessarily a small fraction of  $d$ . Let rays from a source  $P_0$  fall near a point  $P(\xi, \eta, d)$  on its upper face, and be observed at  $Q$  (fig. 2). Let the co-ordinates of  $P_0$  relative to  $P$  be  $(-r_0 \cos \phi, 0, r_0 \sin \phi)$  and those of  $P$  relative to  $Q$

$$(r \sin \psi \cos \chi, r \sin \psi \sin \chi, r \cos \psi), \quad (10)$$

where  $r$  is supposed to be much greater than  $d$ . Since  $\lambda$  is small, the polarization of the disk is nearly the same as that of an infinite disk under plane waves.

Let  $dS$  be an element of area of the upper face. The distance from  $Q$  to a point  $(\xi, \eta, \zeta)$  in the sheet is  $r - (d - \zeta) \cos \psi$ . We can reduce the point  $(\xi, \eta, \zeta)$  to  $P$  as follows. In calculating  $\Pi$  at the point  $Q$ , retardation is allowed for by a factor  $\exp(-ikr)$  in the first of equations (9). We



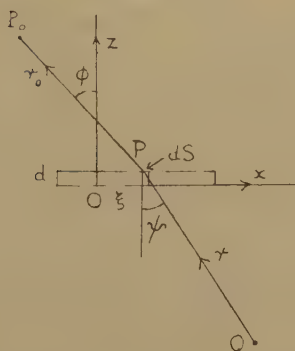
need not alter  $r$  in the denominator, but can convert  $\zeta$  to  $d$  by a factor  $\exp(ikd \cos \psi - ik\zeta \cos \psi)$  in the numerators of  $P_\xi$ ,  $P_\eta$ ,  $P_\zeta$ . Integrating from  $\zeta=0$  to  $d$ , we can flatten the cylinder on  $dS$  into  $dS$  itself. We can thus take the disk to be infinitely thin, and the element  $dS$  to have electric moments  $M_\xi$ ,  $M_\eta$ ,  $M_\zeta$ , where

$$\left. \begin{aligned} [M_\xi] &= -i\epsilon_0(k'^2/k^2 - 1) \cos \phi' dS (AE_3 - BE_5) \\ &\quad \exp(i\omega t - ik\xi \sin \phi - ikr), \\ [M_\eta] &= -i\epsilon_0(k'^2/k^2 - 1) dS (AE_4 + BE_6) \exp(i\omega t - ik\xi \sin \phi - ikr), \\ [M_\zeta] &= -i\epsilon_0(k'^2/k^2 - 1) \sin \phi' dS (AE_3 + BE_5) \\ &\quad \exp(i\omega t - ik\xi \sin \phi - ikr), \end{aligned} \right\} \quad (11)$$

and

$$\left. \begin{aligned} A &= \frac{\exp(ik'd \cos \phi') - \exp(ikd \cos \psi)}{k' \cos \phi' - k \cos \psi}, \\ B &= \frac{\exp(-ik'd \cos \phi') - \exp(ikd \cos \psi)}{-k' \cos \phi' - k \cos \psi}. \end{aligned} \right\} \quad (12)$$

Fig. 2.



The electric intensity of the element  $dS$  at  $Q$  will be called  $dE$ . It is found by applying the formulæ of the Hertzian oscillator to the moment  $M$ , taking only the terms which are greatest at large distances. We have three equations of the type

$$4\pi\epsilon_0 dE_x = \frac{y^2 + z^2}{c^2 r^3} \omega^2 [M_\xi] - \frac{xy}{c^2 r^3} \omega^2 [M_\eta] - \frac{zx}{c^2 r^3} \omega^2 [M_\zeta],$$

where  $(x, y, z)$  are the co-ordinates (10). The incident electric wave at  $P$  is given by equation (1), with  $x$  replaced by  $\xi$  and  $z$  by  $d$ . If it is assimilated to waves of amplitude  $A_1/r_0$  and  $A_2/r_0$  polarized in and perpendicular to the plane of  $xz$ , we must put

$$\left. \begin{aligned} E_1 \exp(-ik\xi \sin \phi + ikd \cos \phi) &= A_1 \exp(-ikr_0)/r_0, \\ E_2 \exp(-ik\xi \sin \phi + ikd \cos \phi) &= A_2 \exp(-ikr_0)/r_0, \end{aligned} \right\} \quad (13)$$

in equation (7).

The effect of the whole disk is found by integration with respect to  $dS$ , in which  $\phi$ ,  $\phi'$ ,  $\psi$ ,  $\chi$  are taken as constant, since their variation is small

compared with that of the exponentials in (11) and (13). We also assume that  $d/r$  and  $d/r_0$  are small enough for  $A$  and  $B$  to be treated as constant. If we write

$$I = \frac{-i \cos \phi}{\lambda r r_0} \int \exp (i \omega t - i k r - i k r_0) dS \quad . \quad . \quad . \quad (14)$$

and

$$\left. \begin{aligned} \Delta_1 Q_x &= (k'^2 - k^2) \cos \phi' [A(k' \cos \phi + k \cos \phi') \\ &\quad + B(k' \cos \phi - k \cos \phi')], \\ \Delta_2 Q_y &= (k'^2 - k^2) [A(k' \cos \phi' + k \cos \phi) + B(k' \cos \phi' - k \cos \phi)], \\ \Delta_1 Q_z &= (k'^2 - k^2) \sin \phi' [A(k' \cos \phi + k \cos \phi') \\ &\quad - B(k' \cos \phi - k \cos \phi')], \end{aligned} \right\} \quad (15)$$

the total electric intensity of the disk at  $Q$  is  $E$ , where

$$\left. \begin{aligned} E_x/IA_1 &= (\sin^2 \psi \sin^2 \chi + \cos^2 \psi) Q_x - \sin^2 \psi \cos \chi \sin \chi Q_y \\ &\quad - \cos \psi \sin \psi \cos \chi Q_z, \\ E_y/IA_2 &= -\sin^2 \psi \cos \chi \sin \chi Q_x + (\sin^2 \psi \cos^2 \chi + \cos^2 \psi) Q_y \\ &\quad - \cos \psi \sin \psi \sin \chi Q_z, \\ E_z/IA_1 &= -\cos \psi \sin \psi \cos \chi Q_x - \cos \psi \sin \psi \sin \chi Q_y + \sin^2 \psi Q_z. \end{aligned} \right\} \quad (16)$$

### 5. Babinet's Principle.

Let a thin infinite homogeneous sheet, of any dielectric constant and conductivity, be cut into two parts  $S_1$  and  $S_2$ , and let a plane wave fall on  $S_1$  and  $S_2$  separately. If the wave-length is small, the polarization at any point of  $S_1$  and  $S_2$  is the same, and equal to that in the infinite sheet. The vector sum of the electric intensities of  $S_1$  and  $S_2$  at any point is that of the infinite sheet, which is zero except on the line of the incident wave. Hence, apart from the central spot,  $S_1$  and  $S_2$  have the same diffraction pattern whatever their size or the material of the sheet. Thus Babinet's principle admits of considerable extension, and is a natural consequence of the smallness of the wave-length. In diffraction by a hole, the total electric intensity is the vector difference of the intensity transmitted by the infinite sheet and that due to the closing disk. We can therefore consider always a disk instead of a perforated sheet, which equation (14) shows to be more convenient.

### 6. Highly Conducting Disk.

If  $\gamma$  is large, the real part of  $ik'$  is large and positive from equations (2) and (3), and  $\cos \phi' \approx 1$ . From (5) and (12),

$$A \approx \exp(ik'd)/k', \quad \Delta_1 \approx k'^2 \cos^2 \phi \exp(ik'd), \quad \Delta_2 \approx k'^2 \exp(ik'd),$$

and  $B$  is small in comparison with  $A$ . Hence from (15)

$$Q_x \approx 1/\cos \phi, \quad Q_y \approx 1, \quad Q_z \approx 0, \quad . \quad . \quad . \quad . \quad . \quad (17)$$

and (16) gives

$$\left. \begin{aligned} E_x/IA_1 &= (\sin^2 \psi \sin^2 \chi + \cos^2 \psi)/\cos \phi - \sin^2 \psi \cos \chi \sin \chi, \\ E_y/IA_2 &= -\sin^2 \psi \cos \chi \sin \chi/\cos \phi + \sin^2 \psi \cos^2 \chi + \cos^2 \psi, \\ E_z/IA_1 &= -\cos \psi \sin \psi \cos \chi/\cos \phi - \cos \psi \sin \psi \sin \chi. \end{aligned} \right\} \quad . \quad (18)$$



For diffraction by a small disk with normal direct vision ( $\phi=\psi=0$ ) we have from (14), (15), and (16)

$$\mathbf{E}/\mathbf{I}\mathbf{A}=1-0.135 \exp(-1070i). \quad . . . . . (22)$$

The diffraction of such a disk on one side is therefore not much different from that of a perfect reflector.

### 8. Transparent Disk.

Take

$$d=10^{-4}, \epsilon=2.25, \gamma=\eta=0, k=1.07 \times 10^7, k'=1.60 \times 10^7, \dots (23)$$

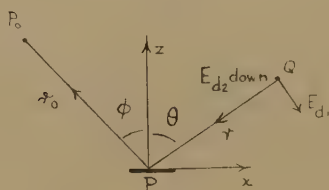
where the value of  $\epsilon$  is the square of the refractive index of common glass. We find

$$\mathbf{E}/\mathbf{I}\mathbf{A}=1-4.84 \exp(1070i)/[5.04 \exp(1600i)-0.20 \exp(-1600i)]. (24)$$

Equations (22) and (24) are included in the general formula

$$\mathbf{E}/\mathbf{I}\mathbf{A}=1-4kk' \exp(ikd)/[(k'+k)^2 \exp(ik'd)-(k'-k)^2 \exp(-ik'd)]. (25)$$

Fig. 3.



The oscillatory factors become more and more prominent as  $\gamma$  is reduced, but the electric intensity is always of the same order of magnitude, if the thickness is not inordinately small.

### 9. Diffraction Intensities.

From (14), (15) and (16), the diffracted wave varies in intensity and polarization from point to point in a way that depends on its shape and on  $\phi$  and  $\psi$ , but the absolute intensity depends also on the polarization and on  $d$ ,  $\epsilon$  and  $\gamma$ . Thus, if the incident light is parallel, the diffraction pattern of a small hole or disk (apart from the central spot) is the same, whatever the polarization or the material of the screen. This simple relation does not hold in Fresnel's more general diffraction phenomena, which depend on interference between the direct and diffracted beams.

The diffraction pattern of a small hole or disk is viewed direct ( $\psi=\phi$ ,  $\chi=\pi$ ), but it is not so for a grating. A grating has a different  $I$  with the property of being large at certain angles  $\psi$  not near to  $\phi$ . Garside\* calculated  $I$  for a concave grating. We can illustrate the theory by finding the effect of polarization on the intensity of grating spectra. The theory will not apply to gratings of more than about 1000 lines to

\* F. A. Garside, Quart. Journ. Math. vii. p. 158 (1936).



the centimetre. Let light fall at angle  $\phi$  on an ideal reflecting grating of equidistant perfectly conducting segments (fig. 3), and consider the radiation received at angle  $\theta$ . Since  $\psi=\pi-\theta$  and  $\chi=\pi$ , equations (18) give

$$E_x/IA_1=\cos^2\theta/\cos\phi, \quad E_y/IA_2=1, \quad E_z/IA_1=-\cos\theta\sin\theta/\cos\phi.$$

Hence, with the notation of the figure,

$$E_{d1}/IA_1=\cos\theta/\cos\phi, \quad E_{d2}/IA_2=1. \quad . \quad . \quad . \quad (26)$$

To fix ideas, suppose that the incident light is originally unpolarized and that a polaroid disk is turned round in front of it so as to make  $A_1=A_2$  in the two positions (26). Then the spectrum is more intense when polarized in the plane of incidence if  $\theta<\phi$ , and less intense if  $\theta>\phi$ .

# XXIX. *On a Relation between some Natural Constants and the Lesser Particles.*

*To the Editors of the Philosophical Magazine.*

GENTLEMEN,—

IN a recent paper Dr. H. D. K. Drew<sup>(1)</sup> shows that the quantity

$$m_s=\frac{2R_\infty h}{c}$$

is nearly equal to

$$\frac{m_e^2}{10M_N},$$

where  $m_e$  and  $M_N$  are the masses of the electron and the neutron respectively. He suggests that the mass  $m_s$  may have some fundamental significance.

This relation is equivalent to the statement that

$$\frac{m_e}{10M_N}=\left(\frac{2\pi e^2}{hc}\right)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (A)$$

Now, if we substitute in this relation the values of  $m_e$  and  $\frac{2\pi e^2}{hc}$ , which are regarded by Birge<sup>(2)</sup> as the most probable values, we obtain

$$\frac{m_e}{10M_N}=54.44\times 10^{-6},$$

while

$$\left(\frac{2\pi e^2}{hc}\right)^2=53.26\times 10^{-6},$$

a difference of about 2 per cent. instead of the agreement to one part in

90,000 claimed by Drew. The disagreement seems to be due mainly to the different values taken for  $m_e$ , namely

$$m_e = 0.0005428835233 \text{ oxygen units (Drew),}$$

$$m_e = 0.00054847 \pm 0.00000017 \text{ oxygen units (Birge).}$$

Granted an approximate agreement between the two expressions for  $m_s$ , it seems rather illogical to suppose the existence of a particle of mass  $m_s$  simply on the grounds that there are two independent ways of calculating its mass from fundamental constants  $\delta$ . Since the arbitrary factor 10 must be included, we might with as much justification write, for instance,

$$m_s = \frac{m_e^2}{M_N} = \frac{20R_\infty h}{c},$$

or multiply each of the two expressions for  $m_s$  by any other numerical factor.

Drew further suggests a picture of the electron as composed of sub-particles of mass  $m_s$ , the mass and kinetic energy of the electron somehow depending upon the arrangement of these sub-particles in such a way as to account for their variation with the electron's velocity. It must be pointed out, however, that the way in which the Special Theory of Relativity explains the observed variation of mass and energy with velocity is so beautifully simple that there can be no case for replacing it by anything else except on grounds of definite experimental evidence. The four-dimensional relativity picture is, in fact, simpler than the Newtonian picture with three dimensions of space and one of time. Moreover, Eddington<sup>(3)</sup> has shown how the ratios

$$\frac{2\pi e^2}{hc} \quad \text{and} \quad \frac{\text{electron mass}}{\text{proton mass}}$$

may be calculated from theoretical considerations starting from relativity and wave mechanics, and although his method is not accepted without reserve by physicists, his results agree remarkably well with experimental values. His results indicate that the relation (A) should be expected to hold approximately, but not exactly. It seems, therefore, quite unnecessary to invoke a particle of mass  $m_s$  in order to account for the approximate equality of two quantities having the dimensions of mass.

Yours faithfully,

(Dr.) A. R. STOKES.

#### *References.*

- (1) Drew, *Phil. Mag.* xxxvi. p. 577 (Aug. 1945).
- (2) Birge, *Rep. Prog. Phys.* viii. pp. 90, 129 (1941).
- (3) Eddington, 'Relativity Theory of Protons and Electrons' (Cambridge University Press, 1936); *Proc. Phys. Soc.* liv. p. 491 (1942).

Royal Holloway College,  
Englefield Green, Surrey.  
July 3, 1946.

[The Editors do not hold themselves responsible for the views  
expressed by their correspondents.]

XXX. *Random Fluctuations in a Cathode Ray Oscillograph.*

By N. R. CAMPBELL, Sc.D., F.Inst.P.\*, and  
V. J. FRANCIS, B.Sc., A.R.C.S., F.Inst.P., A.M.I.E.E.†

(Communication from the Staff of the Research Laboratories of  
The General Electric Company, Limited, England.)

[Received May 11, 1945.]

## SUMMARY.

When a cathode ray oscillograph is preceded by a high gain amplifier, the trace fluctuates irregularly. Such fluctuations are a form of electrical "noise," and arise from the shot and thermal effects that cause noise in valve circuits. This paper applies to them the principles for treating noise set forth in a recent paper<sup>(1)</sup>.

*Sections 1-2.*—The principles are summarized briefly.

*Section 3.*—Certain acknowledgments are made.

*Sections 4-5.*—The chance  $W(y)dy$  is calculated that the deflection of the trace at a snap reading is between  $y$  and  $y+dy$ . If, but only if, the random events giving rise to the noise are sufficiently rapid,  $W(y)$  is Gaussian. The more general case is treated in the Appendix.

*Section 6.*—Certain known theorems about the distribution of correlated quantities are stated.

*Section 7.*—The correlation coefficients are expressed in terms of the  $s(t)$  that represent the effect of a single event.

*Section 8.*—The mean frequency with which the trace cuts a given level is calculated.

*Section 9.*—An alternative method (Bondi's method) of treating the problem of Section 8 is explained.

*Section 10.*—The mean frequency of "peaks" and "troughs" at a given level is calculated.

*Sections 11-15.*—The problem is discussed of determining the mean interval between an upward cut at a given level and the next subsequent downward cut, *i. e.*, the mean duration of a "hill" above this level. The problem is not solved owing to analytical difficulties which are explained.

*Section 16.*—Certain approximate methods of solving the problem of Sections 11-15, and of other similarly intractable problems, are proposed, but no attempt is made to apply them in detail.

*Sections 17-18.*—It is shown how other important information may be

---

\* Late Member of the Research Staff of The General Electric Company, Limited.

† Communicated by C. C. Paterson, F.R.S.





(1.3), (1.4), involve the propositions that, if events are occurring randomly in time as defined, and  $n$  of them happen in an interval  $t$ , then

[illegible]

$$\overline{(n-\bar{n})^2} = \lambda t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.7)$$

## Section 2.

In the previous paper the application of these principles was confined to those problems that are historically the earliest and, perhaps, still of the greatest practical importance, namely the calculation of mean square noise in linear circuits. However, there is another class of problem connected with the trace on the fluorescent screen of a cathode-ray oscillograph. If the oscillograph is preceded by an amplifier of high gain, the trace fluctuates irregularly about the base line when there is no signal; when there is a signal, these fluctuations are imposed on the regular trace due to it and, if the signal is small, may mask it.

If the amplifier and the oscillograph are linear, questions about the mean position of the trace or about its mean square deviation from its mean are all answerable in principle by the methods already described ; and it is no more difficult to determine the essential function  $s(t)$  than it is with other circuits of equal complexity ; the only difference is that  $y$ , instead of being a voltage or current, is the position of the cathode-ray spot on the screen, which is determined by a voltage or current. But other questions may be asked ; for example, what is the probability that the deflection of the ray will have an assigned value at a "snap" observation, how many times a second on the average does the trace cut the base or some line parallel to it, what will be the mean frequency of a peak in the trace, or what the mean interval between peaks or between cuts ?

The primary purpose of this paper is to show that the same principles provide a physically perspicuous method of discussing such questions and of answering some, at least, of them. Others we shall fail to answer because the mathematical problems to which the method leads appear to be insoluble. It is possible that these questions, as well as some of those that we can solve, are better approached from some other standpoint from which it may prove natural to introduce different and, perhaps, sometimes more powerful analytical apparatus; but, even if that is so, it is unlikely that any alternative method will be as perspicuous physically; our presentation will, therefore, still have its value\*.

### Section 3.

Before we proceed, some acknowledgments must be made. Nearly all the results presented here were contained in a report by the present authors circulated in March, 1944, to those interested in this field. While

\* We have in mind here S. O. Rice's paper, Bell System Technical Journal, xxiii. p. 282 (1944), which came to our notice while we were preparing this paper.



except when it is far from  $\bar{n}$ . Then the calculation of  $W(y)$  becomes much simpler.

For it is well known that, if the condition of the special case is fulfilled, the Poisson distribution (4.3) becomes effectively a Gaussian distribution; that is to say, if  $P(n)dn$  is the chance that the number of events occurring in an interval  $\tau$  lies between  $n$  and  $n+dn$ ,  $P(n)$  is the Gaussian function  $G_1(n, \bar{n}, \sigma)$ , defined by

$$G_1(n, \bar{n}, \sigma) \equiv \frac{1}{\sigma\sqrt{2\pi}} e^{-(n-\bar{n})^2/2\sigma^2}, \quad \left. \begin{array}{l} \sigma^2 \equiv (n-\bar{n})^2, \end{array} \right\} \dots \dots \dots (5.1)$$

$\sigma$  in this case being equal to  $\bar{n}$ .  $\sigma$  is the "standard deviation."

Further, it is well known that the distribution of any variable that is a linear function of variables, each of which has a Gaussian distribution, is also Gaussian\*; that is to say, if

$$Z = \Sigma a_r x_r, \quad \dots \dots \dots (5.2)$$

and the chance that any  $x_r$  lies between  $x_r$  and  $x_r+dx_r$  is  $G_1(x_r, \bar{x}_r, \sigma_r)$ , then the chance that  $z$  lies between  $z$  and  $z+dz$  is  $G_1(z, \bar{z}, \sigma_z)$ , where

$$\bar{z} = \Sigma a_r \bar{x}_r, \quad \dots \dots \dots (5.3)$$

$$\sigma_z^2 = \Sigma a_r^2 \sigma_r^2. \quad \dots \dots \dots (5.4)$$

Now according to (1.2),  $y$  is a linear function of the  $n_r$ 's; and in the special case we are considering the distribution of each  $n_r$  is Gaussian; so we may write

$$W(y) = G_1(y, \bar{y}, \sigma_y), \quad \dots \dots \dots (5.5)$$

where, from (5.3, 1.6) and (5.4, 1.7),

$$\bar{y} = \lambda \tau \Sigma s(r\tau) \sim \lambda \int_0^\infty s(t) dt, \quad \dots \dots \dots (5.6)$$

$$\sigma_y^2 = \overline{(y-\bar{y})^2} = \lambda \tau \Sigma \{s(r\tau)\}^2 \sim \lambda \int_0^\infty \{s(t)\}^2 dt. \quad \dots \dots \dots (5.7)$$

Accordingly  $W(y)$  is known when  $\lambda$  and  $s(t)$  are known. Incidentally, it may be remarked, that (5.6, 5.7) are Campbell's theorem. The more elaborate proof given in the previous paper is necessary only in the general case when the distributions of  $x_r$  and of  $y$  cannot be assumed to be Gaussian.

It is important to insist that (5.5) is true only if, when  $\tau$  is taken so small that the sums in (5.6, 5.7) may be identified with the integrals,  $\lambda\tau$  is still a very large number. At the opposite extreme, when  $\lambda$  is so small that a second event seldom occurs before the effect of its predecessor

---

\* This proposition is used in the conventional theory of errors for deducing the probable error of a sum from the probable errors of its components. The general proof, though neat, is somewhat abstruse. When few variables are involved, it can easily be proved directly by substituting for one of the  $x$ 's in terms of  $z$  and integrating over all values of the other  $x$ 's, which are now independent variables.

has died away, (5.5) cannot be true, except possibly for a very special form of  $s(t)$ . Proofs that  $y$  has a Gaussian distribution which do not involve this condition must be fallacious\*. But in the sequel we shall assume a Gaussian distribution.

### Section 6.

The rest of the questions suggested in Section 2 involve, not an isolated value of  $y$ , but two or more values of  $y$  in prescribed time relations. Thus the chance that the trace will cut the base line in an upward direction in some specified interval is the chance that, within that interval, a negative value of  $y$  will be followed almost immediately by a positive value. Now the chance  $W(y_1, y_2)dy_1dy_2$  that  $y$  will be  $y_1$  to  $y_1+dy_1$  at  $t$  and  $y_2$  to  $y_2+dy_2$  at  $t+t_1$  has its "normal" value  $W(y_1)W(y_2)dy_1dy_2$  only if  $t_1$  is so great that the effects of the events that contributed to  $y_1$  have died away before  $y_2$  is observed; for it is only then that both readings can be considered "snap" readings. If  $t_1$  is smaller than this,  $y_1$  and  $y_2$  are correlated, *i. e.*,  $W(y_1, y_2)$  may be greater or less than  $W(y_1)W(y_2)$ . If  $t_1$  is very small, it will clearly be greater if  $y_1$  and  $y_2$  are nearly the same, and less if they are very different.

The necessary and sufficient conditions for lack of correlation between two quantities  $u$  and  $v$  is that  $W(u, v)$  can be expressed in the form  $W_1(u)W_2(v)$ , where  $W_1(u)$ ,  $W_2(v)$  are independent of  $v$  or  $u$  respectively. A necessary, but not a sufficient, condition is that

$$\overline{(u-\bar{u})(v-\bar{v})}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

On the other hand, if  $u$  and  $v$  differ only by a constant factor, so that  $v=\beta u$ ,

$$\overline{(u-\bar{u})(v-\bar{v})}=\beta\overline{(u-\bar{u})^2}=\beta\sigma_u^2=\frac{1}{\beta}\sigma_v^2, \quad . \quad . \quad . \quad . \quad . \quad (6.2)$$

where  $\sigma_u^2$ ,  $\sigma_v^2$  are the variances of  $u$  and  $v$ . Accordingly the quantity  $\rho_{uv}$ , defined by

$$\rho_{uv} \cdot \sigma_u \sigma_v = \overline{(u-\bar{u})(v-\bar{v})}, \quad . \quad . \quad . \quad . \quad . \quad (6.3)$$

is zero when  $u$  and  $v$  are uncorrelated and 1 when they are most closely connected. It is therefore a fit measure of correlation and is called the "correlation coefficient."  $\rho_{vv}$  or  $\rho_{uu}$  is always 1.

We can now proceed to state a known proposition of probability theory that provides answers to some at least of the questions we have in view; the proposition stated in Section 5 is merely a very simple case of it. Let each of the  $n$  variables  $z_1 \dots z_l \dots z_n$  be a linear function of a set of variables  $x_1 \dots x_r$ , so that

$$z_1 = \sum_r a_{1r} x_r; \quad \dots; \quad z_l = \sum_r a_{lr} x_r. \quad . \quad . \quad . \quad . \quad . \quad (6.4)$$

Suppose that no pair of the  $x_r$ 's is correlated and that the distribution

---

\* *E. q.*, V. D. Landon, *Proc. I. R. E.* xxix, p. 50 (1941).



of each of them is Gaussian. Then if  $W(z_1 \dots z_n) dz_1 \dots dz_n$  is the chance that  $z_1$  lies between  $z_1$  and  $z_1 + dz_1$ ,  $z_l$  between  $z_l$  and  $z_l + dz_l$

$$W(z_1 \dots z_l \dots z_n)$$

can be expressed as a function of the  $z_l$  of their means  $\bar{z}_l$  of their standard deviations  $\sigma_l$ , and their correlation coefficients in pairs  $\rho_{kl}$

$$W(z_1 \dots z_n) dz_1 \dots dz_n = \frac{1}{(2\pi)^{n/2} \sigma_1 \dots \sigma_n \Delta^{1/2}} \times \exp \left[ -\frac{1}{2\Delta} \sum_{kl} \frac{\Delta_{kl} (z_k - \bar{z}_k)(z_l - \bar{z}_l)}{\sigma_k \sigma_l} \right] dz_1 \dots dz_n, \quad (6.5)$$

where  $\Delta$  is the determinant.

$$\Delta = \begin{vmatrix} \rho_{11} & \dots & \rho_{1l} & \dots & \rho_{1n} \\ \vdots & & \vdots & & \vdots \\ \rho_{k1} & \dots & \rho_{kl} & \dots & \rho_{kn} \\ \vdots & & \vdots & & \vdots \\ \rho_{n1} & \dots & \rho_{nl} & \dots & \rho_{nn} \end{vmatrix}, \quad \dots \dots \dots (6.6)$$

and  $\Delta_{kl}$  is the cofactor of  $\rho_{kl}$  in  $\Delta$ . The  $\sum_{kl}$  is taken over all pairs  $(k, l)$ ; but since  $\rho_{kl} = \rho_{lk}$ , we may omit one of the terms  $(k, l)$  and  $(l, k)$  where  $k \neq l$ , and introduce a factor 2 before the other. When there are  $n$   $z$ 's, the expression on the right of (6.5) will be denoted by  $G_n$ ; the values of  $G_2$  and  $G_3$  are given in full in Sections 8, 10 below,  $\rho_{kk}$  being always replaced by 1.

### Section 7.

The correlation coefficients have now to be calculated for the cases that concern us. From henceforward until Section 17 we shall assume that there is no "signal," and that the deflection  $y$  is due wholly to noise representing the superposed effects of events random in time. Further, we shall assume that there is one set only of such events; for if there are two or more sources of noise, *e. g.* two resistors or two valves, the effects that they produce are uncorrelated and the chance that two quantities occur, one the resultant of one set and the other of the other, is simply the product of the chances that either resultant occurs separately.

Suppose that an event produces two effects, represented by  $s_1(t)$  and  $s_2(t)$ , that  $y_1$  is the resultant of the superposed effects  $s_1$ , and  $y_2$  the resultant, observed simultaneously with  $y_1$  of the superposed effects  $s_2$ . Then, it follows from (6.3) and (1.5) that  $\rho_{12}$ , the correlation coefficient of  $y_1$  and  $y_2$ , is given by

$$\sigma_1 \sigma_2 \rho_{12} = \lambda \int_0^\infty s_1(t) s_2(t) dt, \quad \dots \dots \dots (7.1)$$

where

$$\sigma_1^2 = \lambda \int_0^\infty \{s_1(t)\}^2 dt; \quad \sigma_2^2 = \lambda \int_0^\infty \{s_2(t)\}^2 dt. \quad \dots \dots (7.2)$$

(7.1) is true only if the resultants  $y_1$  and  $y_2$  are observed simultaneously. In certain cases of great importance for our purpose, one of the correlated resultants ( $y_2$ ) is observed at a fixed interval  $t_1$  after the observation of the other ( $y_1$ ). In order to find the correlation coefficients in this case, it is best to return to the fundamental equation (1.2). Since the intervals  $\tau$  are counted backwards from the instant of observation we must now count them from the later observation  $y_2$ . If, then, we write  $m\tau = t_1$ , and (for brevity)  $s_{1r}, s_{2r} \dots$  for  $s_1(r\tau), s_2(r\tau) \dots$ , we have in analogy with (1.2)

$$\left. \begin{aligned} y_2 &= n_1 s_{21} + n_2 s_{22} + \dots n_{m+1} s_{2, m+1} + \dots n_{m+N} s_{2, m+N}, \\ y_1 &= n_1 0 + n_2 0 + \dots n_{m+1} s_{11} \dots n_{m+N} s_{1N} \end{aligned} \right\} \dots (7.3)$$

Proceeding by a slight modification of the argument by which (1.4) was proved in the previous paper,

$$(y_1 - \bar{y}_1)(y_2 - \bar{y}_2) = \sum_{r=1}^N \sum_{q=1-m}^N (n_{m+r} - \bar{n}_{m+r}) s_{1r} \cdot (n_{m+q} - \bar{n}_{m+q}) s_{2, m+q} \dots (7.4)$$

Since the numbers of events occurring in different intervals are uncorrelated,

$$\overline{(n_{m+r} - \bar{n}_{m+r})(n_{m+q} - \bar{n}_{m+q})} = 0 \quad \text{if } r \neq q. \dots (7.5)$$

Further when  $r = q$ , by (1.7),

$$\overline{(n_{m+r} - \bar{n}_{m+r})^2} = \lambda \tau. \dots (7.6)$$

Hence

$$\left. \begin{aligned} \overline{(y_1 - \bar{y}_1)(y_2 - \bar{y}_2)} &= \lambda \tau \sum_r s_{1r} s_{2, m+r} \\ \sigma_1 \sigma_2 \rho_{12} &= \lambda \int_0^\infty s_1(t) s_2(t+t_1) dt \end{aligned} \right\} \dots (7.7)$$

and

Suppose now that  $s_1$  and  $s_2$  are the same, so that the suffixes of  $s$  and  $\sigma$  may be dropped and that  $t_1$  is a very short interval  $\tau$  that is ultimately to be reduced to zero. Then, expanding  $s(t+t_1)$  by Taylor's theorem,

$$\sigma^2 \rho_{12} = \lambda \int_0^\infty \left[ \{s(t)\}^2 + \tau s(t) \dot{s}(t) + \frac{\tau^2}{2} s(t) \ddot{s}(t) \dots \right] dt. \dots (7.8)$$

Now

$$\int_0^\infty s(t) \dot{s}(t) dt = \frac{1}{2} \int_0^\infty \frac{d}{dt} \{s(t)\}^2 dt = \frac{1}{2} [\{s(\infty)\}^2 - \{s(0)\}^2]. \dots (7.9)$$

$s(\infty)$  is zero in virtue of one of the fundamental assumptions of Section 1. In practice  $s(0)$  is also zero, because no event can produce an effect that is truly instantaneous. Consequently the second term of the integrand in (7.8) vanishes. Similarly, we may show that the third term of the integrand is

$$-\frac{\tau^2}{2} \int_0^\infty \{\dot{s}(t)\}^2 dt. \dots (7.10)$$



reading is infinitesimal;  $W(y_0)$  enters only in the form  $W(y_0)dy$ , which is the chance that  $y$  lies between  $y_0$  and  $y_0+dy$ . This chance is not of the form we need; we need a chance of the form  $C\tau$ , where  $\tau$  is an infinitesimal interval; for it is only then that we can identify  $C$  with the mean frequency of a cut\*. We must direct our calculations so as to obtain a chance of this form, starting from chances of the form  $W(z_1, z_2 \dots)dz_1 dz_2 \dots$ . In the calculation to which we proceed, the transition is made by eliminating the  $dz$ 's by integration, and introducing  $\tau$  through the correlation coefficients.

If there is an upward cut at level  $y_0$  between  $t$  and  $t+\tau$ , where  $\tau$  is infinitesimal,  $y_1$ , the value of  $y$  at  $t$ , must lie between  $-\infty$  and  $y_0$ , and  $y_2$ , the value of  $y$  at  $t+\tau$ , must lie between  $y_0$  and  $+\infty$ . If, anticipating our result, we write the chance of the cut  $C(y_0) \cdot \tau$ , we must have, in virtue of Section 6,

$$C(y_0) \cdot \tau = \int_{-\infty}^{y_0} dy_1 \int_{y_0}^{+\infty} dy_2 G_2(y_1, y_2), \quad \dots \quad (8.1)$$

where

$$G_2 = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho_{12}^2)^{1/2}} \exp \left[ -\frac{1}{2(1-\rho_{12}^2)} \left\{ \frac{(y_1 - \bar{y}_1)^2}{\sigma_1^2} - \frac{2\rho_{12}(y_1 - \bar{y}_1)(y_2 - \bar{y}_2)}{\sigma_1\sigma_2} + \frac{(y_2 - \bar{y}_2)^2}{\sigma_2^2} \right\} \right] \quad \dots \quad (8.2)$$

In our case  $\sigma_1 = \sigma_2 = \sigma$ ; and  $\rho_{12}$  is given by (7.12); further, we suppose that the base line has been made to coincide with the mean position of the trace, so that  $\bar{y}_1 = \bar{y}_2 = 0$ . If we insert these values, and at the same time substitute for the variables  $y_1 y_2$  the variables  $x_1, x_2$ , where

$$x_1 = \frac{y_0 - y_1}{\sigma\tau}; \quad x_2 = \frac{y_2 - y_0}{\sigma\tau} \quad \dots \quad (8.3)$$

We get

$$C(y_0)\tau = \tau \cdot \frac{\sigma}{2\pi\sigma} \int_0^\infty dx_1 \int_0^\infty dx_2 \exp \left[ -\left\{ \frac{x_1 + x_2}{2} \right\}^2 + \frac{y_0^2}{2\sigma^2} + \frac{1}{2\sigma^2} (y_0 \sigma\tau x_2 - x_1 - \sigma^2\tau^2 x_1 x_2) \right] \quad \dots \quad (8.4)$$

Taking the part of the exponential independent of  $x_1, x_2$  outside the integral, and expanding the part of it involving  $\tau$ , we have

---

\* The only justification that has been given so far for identifying  $C$  with the mean frequency when the chance is  $C\tau$  is (1.6). This equation holds only when the events are random in time. Formally, therefore, we should prove either that upward cuts are random in time according to the definition of Section 1, or that the identification is permissible more generally. Both of these propositions are true, though there may be some limitations on the second; we shall not offer a formal proof of either.



$$C(y_0) = \frac{\sigma^{(1)}}{2\pi\sigma} e^{-\frac{y_0^2}{2\sigma^2}} \int_0^\infty dx_1 \int_0^\infty dx_2 \exp\left\{-\frac{(x_1+x_2)^2}{2}\right\} \left[1 + \frac{1}{2\sigma^2} (y_0 \sigma \tau x_2 - x_1 - \sigma^2 \tau^2 x_1 x_2) + \frac{1}{2!} \left\{ \frac{1}{2\sigma^2} (y_0 \sigma \tau x_2 - x_1 - \sigma^2 \tau^2 x_1 x_2) \right\}^2 \right] \quad (8.5)$$

$$\begin{aligned} &= \frac{\sigma^{(1)}}{2\pi\sigma} \cdot e^{-\frac{y_0^2}{2\sigma^2}} \left[ \int_0^\infty dx_1 \int_0^\infty dx_2 \exp\left\{-\frac{(x_1+x_2)^2}{2}\right\} \right. \\ &\quad + \tau \frac{\sigma y_0}{2\sigma^2} \int_0^\infty dx_1 \int_0^\infty dx_2 (x_2 - x_1) \exp\left\{-\frac{(x_1+x_2)^2}{2}\right\} \\ &\quad + \tau^2 \frac{\sigma^2}{2\sigma^2} \int_0^\infty dx_1 \int_0^\infty dx_2 \left\{ -x_1 x_2 + \frac{y_0^2 (x_2 - x_1)^2}{4\sigma^2} \right\} \\ &\quad \left. \times \exp\left\{-\frac{(x_1+x_2)^2}{2}\right\} \right] \quad (8.6) \end{aligned}$$

Each of the double integrals is finite; accordingly, in the limit  $\tau \rightarrow 0$ , all terms in the outside brackets except the first vanish.

The integrals extend over a quadrant of the  $(x_1, x_2)$  plane. If this plane is divided into strips of width  $d\eta$  by lines  $x_1 + x_2 = \eta$ , the area of each strip is  $\eta d\eta$ , and its contribution to the double integrals is

$$\eta d\eta \cdot e^{-\frac{\eta^2}{2}} \quad (8.7)$$

Hence this double integral may be replaced by the single integral

$$\int_0^\infty \eta e^{-\frac{\eta^2}{2}} d\eta = 1, \quad (8.8)$$

and we obtain finally

$$C(y_0) = \frac{\sigma^{(1)}}{2\pi\sigma} e^{-\frac{y_0^2}{2\sigma^2}} \quad (8.9)$$

This is the mean frequency of upward cuts at the level  $y_0$ , a trivial modification of the calculations gives the mean frequency of downwards cuts; it turns out again to be  $C(y_0)$  as it should, since upward and downward cuts must occur in pairs.

### Section 9.

There is another method (Bondi's method) of solving the problem of Section 8 that is mathematically simpler, but involves conceptions of some importance that have not yet been introduced. An upward cut at  $y_0$  can also be defined as the combination of (1) a value  $y_0$  for  $y$ , and (2) the occurrence of a positive value for  $\dot{y}$ , the velocity of the trace perpendicular to the base. This velocity is the resultant of the velocities derived from the individual events, just as the position of the trace is the resultant of the positions derived from those events. But the velocity derived from a single event is represented by  $\dot{s}(t)$ ; and it is easy to see

that, if the effects of individual events represented by  $s(t)$  are additive, so are those represented by  $\dot{s}(t)$ .  $\dot{y}$  is related to  $\dot{s}(t)$  exactly as  $y$  is related to  $s(t)$ ;  $\dot{y}$  is simply another effect of the same events, the effect being now represented by  $\dot{s}(t)$ , instead of  $s(t)$ . Corresponding propositions are true, if there is substituted for  $\dot{y}$  and  $\dot{s}(t)$  any higher derivatives  $y^{(n)}$  and  $s^{(n)}(t)$ .

Since the event (1) is now a value of  $y$  and the event (2) a value of  $\dot{y}$ , and since  $\dot{y}$ , as well as  $y$ , is a linear function of the  $n_r$  in (1.2), we may apply (7.1) and write for the correlation coefficient

$$\sigma \sigma \rho_{12} = \lambda \int_0^\infty s(t) \dot{s}(t) dt, \quad \dots \dots \dots (9.1)$$

where  $\sigma$ , the standard deviation of  $\dot{y}$ , is the quantity already so denoted and defined by (7.11). But  $\rho_{12}$  is zero in virtue of (7.9). Hence the chance that, at time  $t-\tau$ ,  $y$  lies between  $y_0$  and  $y_0+dy_0$ , while  $\dot{y}$  is positive, *i. e.* between 0 and  $+\infty$ , is

$$\frac{1}{2\pi\sigma\sigma} \int_{\dot{y}=0}^\infty \exp \left[ -\frac{1}{2} \left\{ \frac{y_0^2}{\sigma^2} + \frac{\dot{y}^2}{\sigma^2} \right\} \right] dy_0 d\dot{y}. \quad \dots \dots \dots (9.2)$$

In order to obtain the chance we need, we must impose the further limitation that the trace reaches  $y_0$  not later than  $t$ . Now if, at  $t-\tau$ ,  $y$  is  $y_0-y_1$  and  $\dot{y}$  is the upward velocity of the trace, the trace cannot reach  $y_0$  at or before  $t$  unless  $y_1 \geq \dot{y}\tau$ . But  $dy_0$  in (9.2) is the greatest departure of  $y$  from  $y_0$  at  $t-\tau$ ; hence this further limitation is imposed by writing  $dy_0 = \dot{y}\tau$ , and (9.2) leads to

$$C(y_0)\tau = \frac{\tau}{2\pi\sigma\sigma} e^{-\frac{y_0^2}{2\sigma^2}} \int_0^\infty \dot{y} e^{-\frac{\dot{y}^2}{2\sigma^2}} d\dot{y}, \quad \dots \dots \dots (9.3)$$

that is

$$C(y_0) = \frac{\sigma}{2\pi\sigma} e^{-\frac{y_0^2}{2\sigma^2}}, \quad \dots \dots \dots (9.4)$$

agreeing with (8.9).

### Section 10.

Similar arguments enable us to find  $P_1(y_0)dy_0$ , the mean frequency of peaks (see fig. 1) at a level between  $y_0$  and  $y_0+dy_0$ . (We must now include the factor  $dy_0$ , for the chance of a peak at exactly a defined level is infinitesimal.) We shall give only the argument according to Bondi's method.

We have now three events :—(1)  $y$  is between  $y_0$  and  $y_0+dy_0$  at  $t-\tau$ ; (2)  $y$  is between 0 and  $+\dot{y}$ ; (3)  $\dot{y}$  is negative, *i. e.* between 0 and  $-\infty$ , subject to the condition (analogous to  $dy_0 = \dot{y}\tau$  in the previous argument) that  $dy = \dot{y}\tau$ . (These conditions do not exclude a possibility that the

peak occurs *just* outside the range  $dy_0$ , but it is easy to see that any error on this score is infinitesimal.) We have, therefore,

$$P_1(y_0)dy_0 = \int_{y_0}^{y_0+dy_0} dy_0 \int_0^{dy} dy \int_{-\infty}^0 G_3 d\dot{y}. \quad (10.1)$$

$G_3$ , written in full, is

$$\frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3\Delta^{1/2}} \cdot \exp \left[ -\frac{1}{2\Delta} \left\{ \frac{z_1^2(1-\rho_{23}^2)}{\sigma_1^2} + \frac{z_2^2(1-\rho_{13}^2)}{\sigma_2^2} + \frac{z_3^2(1-\rho_{12}^2)}{\sigma_3^2} \right. \right. \\ \left. \left. - \frac{2z_1z_2(\rho_{12}-\rho_{13}\rho_{23})}{\sigma_1\sigma_2} - \frac{2z_2z_3(\rho_{23}-\rho_{12}\rho_{13})}{\sigma_2\sigma_3} - \frac{2z_1z_3(\rho_{13}-\rho_{12}\rho_{23})}{\sigma_1\sigma_3} \right\} \right], \quad (10.2)$$

with

$$\Delta = 1 - \rho_{12}^2 - \rho_{23}^2 - \rho_{13}^2 + 2\rho_{12}\rho_{23}\rho_{13}.$$

We identify  $z_1, z_2, z_3$  respectively with  $y_0, y$  (which is 0),  $\dot{y}$ . Then  $\sigma_1, \sigma_2$  are defined respectively by (5.7) and (7.11), and we write

$$\left. \begin{aligned} \sigma_1 &= \sigma \\ \sigma_2 &= \sigma \\ \sigma_3^2 &= \sigma^2 = \lambda \int_0^\infty \{\dot{s}(t)\}^2 dt \end{aligned} \right\}, \quad (10.3)$$

by (7.9)

$$\sigma \sigma \rho_{12} = \lambda \int_0^\infty s(t)\dot{s}(t) dt = 0, \quad (10.4)$$

by (7.10)

$$\sigma \sigma \rho_{13} = \lambda \int_0^\infty s(t)\ddot{s}(t) dt = -\sigma^2, \quad (10.5)$$

$$\begin{aligned} \sigma \sigma \rho_{23} &= \lambda \int_0^\infty \dot{s}(t)\ddot{s}(t) dt \\ &= \lambda \int_0^\infty \frac{1}{2} \frac{d}{dt} \{\dot{s}(t)\}^2 dt = \frac{\lambda}{2} \{\dot{s}(0)\}^2. \end{aligned} \quad (10.6)$$

Then

$$P_1(y_0) = \frac{1}{(2\pi)^{3/2}\sigma \sigma \sigma \Delta^{1/2}} \int_{-\infty}^0 \exp \left[ -\frac{1}{2\Delta} \left\{ \frac{y_0^2(1-\rho_{23}^2)}{\sigma^2} + \frac{\dot{y}^2}{\sigma^2} \right. \right. \\ \left. \left. - \frac{2y_0\dot{y}\rho_{13}}{\sigma \sigma} \right\} \right] \dot{y} d\dot{y} \quad (10.7)$$

$$= \frac{\Delta^{1/2}\sigma^{(2)}}{(2\pi)^{3/2}\sigma \sigma} e^{-\frac{y_0^2}{2\sigma^2}} \cdot E_1 \left( \frac{y_0}{\sigma} \cdot \frac{\rho_{13}}{\sqrt{2\Delta}} \right), \quad (10.8)$$

where

$$E_1(x) = e^{-x^2} - \sqrt{\pi}x(1 - \text{Erf } x). \quad (10.9)$$

The second term on the right of (10.9) is never greater than the first but tends to equality with it as  $x \rightarrow \infty$ ; consequently  $E_1(-x)/E_1(+x)$ ,

increases from 1 to  $\infty$  as  $x$  increases from 0. Now  $\rho_{13}$  must be negative by (10.5). Consequently  $P_1(y_0)$  is greater when  $y_0$  is positive than when it is negative, and peaks at a given distance from the base are more frequent when they are above than when they are below the base.

If  $P_2(y_0)dy_0$  is the mean frequency of a trough (fig. 1) at level between  $y_0$  and  $y_0+dy_0$ , we have to replace conditions (2) and (3) by (2')  $\dot{y}$  is between 0 and  $-\dot{d}y$ , a negative quantity, (3')  $y$  is between 0 and  $+\infty$ , subject to  $\dot{d}y=\ddot{y}\tau$ . In (10.1) we then change the sign of the integrand and replace the upper and lower limits of integration by  $+\infty$  and 0 respectively. We then get

$$P_2(y_0) = \frac{\Delta^{1/2} \sigma^{(2)}}{(2\pi)^{3/2} \sigma^{(1)}} e^{-\frac{y_0^2}{2\sigma^2}} E_2\left(\frac{y_0}{\sigma} \cdot \frac{\rho_{13}}{\sqrt{2\Delta}}\right), \quad \dots \quad (10.10)$$

where

$$E_2(x) = e^{-x^2} + \sqrt{\pi}x(1 + \text{Erf } x), \quad \dots \quad (10.11)$$

so that troughs at a given distance from the base are more frequent when they are below than when they are above it. If the trace is optically reflected in the base line, peaks become troughs, "above" becomes "below," and *vice versa*. Accordingly, so far as the mean frequency of peaks and troughs are concerned, since  $E_1(-x) = E_2(x)$ , the trace is symmetrical about the base.

From (10.8) and (10.10) it follows that the mean frequency of occurrence of turning points on the trace (peaks and troughs) is

$$P_1(y_0) + P_2(y_0) = \frac{2\Delta^{1/2} \sigma^{(2)}}{(2\pi)^{3/2} \sigma^{(1)}} \int_{-\infty}^{+\infty} e^{-\frac{y_0^2}{2\sigma^2}} \left\{ e^{-\frac{y_0^2 \rho_{13}^2}{2\sigma^4 \Delta}} + \frac{y_0 \rho_{13}}{\sigma} \sqrt{\frac{\pi}{2\Delta}} \text{Erf}\left(\frac{y_0 \rho_{13}}{\sigma \sqrt{2\Delta}}\right) \right\} dy_0, \quad \dots \quad (10.12)$$

and by actual integration it is not difficult to show that this gives

$$P_1(y_0) + P_2(y_0) = \frac{\sigma^{(2)} (1 - \rho_{23}^2)^{1/2}}{\pi \sigma^{(1)}}. \quad \dots \quad (10.13)$$

Now, if we start with a trace derived from events  $s(t)$  and differentiate it, a trough or a peak will become a cut on the base line. Our differentiated trace can be considered as derived from events  $\dot{s}(t)$ , in which case (9.4) shows that the mean frequency of cuts at the base line should be

$$\frac{\sigma^{(2)}}{\pi \sigma^{(1)}}. \quad \dots \quad (10.14)$$

This agrees with (10.13) as it should because in deriving (9.4) we assumed that  $s(0) = 0$ ; so that in order to satisfy a similar condition for our differentiated curve we must have  $\dot{s}(0) = 0$ , which gives  $\rho_{23} = 0$  in (10.13).



## Section 11.

The last of the questions asked in Section 2, namely, the mean interval between peaks or between cuts, involves the correlation between events that are separated by finite intervals. This introduces no new questions of principle if the method of Section 8 is adopted; but it does if Bondi's method is adopted throughout. If the trace were representable by a regular mathematical function  $y=f(t)$ , so that the future of the trace could be predicted from a complete knowledge of its past, the condition that the trace, having cut the level  $y_0$  at  $t=t_0$  would cut it again at  $t=t_2$ , could be stated in terms of the derivatives of  $y$  at  $t_0$ . An actual trace, whose future course cannot be predicted however complete the knowledge of the past, cannot have derivatives in the mathematical sense that are all defined. The question arises therefore whether, if the derivatives are defined as we have defined them, namely as the resultants of the various derivatives of  $s(t)$ , the chance that the trace will cut the level again at  $t_2$  can be stated in terms of these derivatives. The answer may seem doubtful, because the course of the trace near  $t_2$  is determined by events that happen after  $t_0$  and do not therefore enter into the definition of the derivatives; but it must be remembered that these later events are concerned in the correlation coefficients which take all events into account. Detailed inquiry shows that, in spite of this objection, the answer is affirmative; but we need not discuss the matter, because it turns out that, even if the consistent use of Bondi's method is permissible in principle, the mathematical difficulties to which it leads are even greater than those arising in the method of Section 8. It is only when all the relevant values of the variable and its derivatives are simultaneous that Bondi's method is mathematically the easier.

## Section 12.

We shall therefore adopt a mixed method. Suppose that the question is, what is the chance that, when the trace has made an upward cut of the level  $y_0$  at  $t$ , it will make a downward cut at the same level between  $t+t_2$  and  $t+t_2+\tau_2$ , where  $\tau_2$  is an infinitesimal interval? The chance that the two cuts occur is the chance of the happening of the four events: (1)  $y$  between  $y_0$  and  $y_0-dy_1$  at  $t-\tau_1$ ; (2)  $\dot{y}_1$  between 0 and  $+\infty$  at  $t-\tau_1$ , with  $dy_1=\dot{y}_1\tau_1$ ; (3)  $y$  between  $y_0$  and  $y_0+dy_2$  at  $t+t_2+\tau_2$ ; (4)  $\dot{y}_2$  between 0 and  $-\infty$  at  $t+t_2-\tau_2$ , with  $dy_2=\dot{y}_2\tau_2$ . (Suffixes are attached to the  $\dot{y}$  to indicate that the  $\dot{y}$ 's in (2) and (4) are independent variables.) This chance will be of the form  $C(y_0, t_2)\tau_2\tau_1$ . The chance that the second cut occurs when it is known that the first has occurred is this chance divided by the chance that the first occurs, that is to say it is

$$C_1(y_0t_2)\tau_2=C(y_0, t_2)\tau_2\tau_1/C(y_0)\tau_1, \quad . \quad . \quad . \quad (12.1)$$

where  $C(y_0)$  is given by (8.9).

It is important to notice that  $C_1(y_0, t_2)$  is *not* the mean frequency of a "hill" above  $y_0$  (see fig. 1) of duration  $t_2$ . There are two reasons, one

trivial and one very far from trivial. The trivial reason is that the chance of an exactly prescribed duration  $t_2$  must be zero; the only significant quantity is the chance, proportional to  $\tau_2$  when  $\tau_2$  is small, that the duration lies between  $t_2$  and  $t_2 + \tau_2$ . The other reason is that  $C_1(y_0, t_2)\tau_2$ , given by (12.1), is *not* the chance that the downward cut that must follow the upward cut at  $t$  before any other cut occurs, occurs at  $t + t_2$ . It is the chance that *any* downward cut occurs at  $t + t_2$ ; it does not distinguish between the case in which the next cut (necessarily a downward cut) occurs at  $t + t_2$ , and cases in which one or more downward cuts (each necessarily preceded and followed by an upward cut) occur between  $t$  and  $t + t_2$ .  $C_1(y_0, t_2)$  is not therefore a quantity that by itself is likely to be of much practical importance; nevertheless, we shall calculate it before we proceed.

### Section 13.

If, in distinguishing the various  $\sigma$ 's and  $\rho$ 's, the suffixes 1, 2, 3, 4 denote respectively the events (1), (2), (3), (4) defined above, we have

$$\sigma_1 = \sigma_3 = \sigma; \quad \sigma_2 = \sigma_4 = \overset{(1)}{\sigma}, \quad . \quad . \quad . \quad . \quad . \quad (13.1)$$

where  $\overset{(1)}{\sigma}, \sigma$  have the meanings previously assigned to them.

$$\left. \begin{aligned} \rho_{12} &= \rho_{34} = 0 \\ \sigma^2 \rho_{13} &= \lambda \int_0^\infty s(t)s(t+t_2) dt \\ \overset{(1)}{\sigma} \sigma \rho_{14} &= \lambda \int_0^\infty s(t)\dot{s}(t+t_2) dt \\ \overset{(1)}{\sigma} \sigma \rho_{23} &= \lambda \int_0^\infty \dot{s}(t)s(t+t_2) dt \\ \overset{(1)}{\sigma}^2 \rho_{24} &= \lambda \int_0^\infty \dot{s}(t)\dot{s}(t+t_2) dt \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad (13.2)$$

$$\left. \begin{aligned} \Delta &= 1 - \rho_{13}^2 - \rho_{14}^2 - \rho_{23}^2 - \rho_{24}^2 + \rho_{13}^2 \rho_{24}^2 + \rho_{14}^2 \rho_{23}^2 - 2\rho_{13}\rho_{14}\rho_{23}\rho_{24} \\ \Delta_{11} &= 1 - \rho_{23}^2 - \rho_{24}^2 & \Delta_{12} &= \rho_{13}\rho_{23} + \rho_{14}\rho_{24} & \Delta_{34} &= \rho_{13}\rho_{14} + \rho_{23}\rho_{24} \\ \Delta_{22} &= 1 - \rho_{13}^2 - \rho_{14}^2 & \Delta_{13} &= -\rho_{13}(1 - \rho_{24}^2) - \rho_{14}\rho_{23}\rho_{24} \\ \Delta_{33} &= 1 - \rho_{14}^2 - \rho_{24}^2 & \Delta_{14} &= -\rho_{14}(1 - \rho_{23}^2) - \rho_{13}\rho_{23}\rho_{24} \\ \Delta_{44} &= 1 - \rho_{13}^2 - \rho_{23}^2 & \Delta_{23} &= -\rho_{23}(1 - \rho_{14}^2) - \rho_{13}\rho_{14}\rho_{24} \\ & & \Delta_{24} &= -\rho_{24}(1 - \rho_{13}^2) - \rho_{13}\rho_{14}\rho_{23} \end{aligned} \right\} . \quad (13.3)$$

$$C(y_0, t_2)\tau_1\tau_2 = - \int_0^\infty dy_1 \int_{-\infty}^0 dy_2 G_4(y_0, \dot{y}_1, y_0\dot{y}_2)\dot{y}_1\tau\dot{y}_2\tau \quad . \quad . \quad . \quad . \quad (13.4)$$

$$= \frac{\tau_1\tau_2}{4\pi^2\overset{(1)}{\sigma}^2\sigma^2\Delta^{1/2}} \int_0^\infty dy_1 \int_0^\infty dy_2 \cdot \dot{y}_1\dot{y}_2 \exp \left[ -\frac{1}{2\Delta} U \right], \quad . \quad (13.5)$$

where

$$\begin{aligned} U = & \frac{y_0^2}{\sigma^2} (\Delta_{11} + \Delta_{33} + 2\Delta_{13}) + \frac{(\dot{y}_1)^2}{\frac{(1)}{\sigma_2}} \Delta_{22} + \frac{(\dot{y}_2)^2}{\frac{(1)}{\sigma^2}} \Delta_{44} \\ & + \frac{2y_0}{\frac{(1)}{\sigma\sigma}} \{ \dot{y}_1(\Delta_{12} + \Delta_{23}) - \dot{y}_2(\Delta_{14} + \Delta_{34}) \} - \frac{2\dot{y}_1\dot{y}_2}{\sigma^2} \Delta_{24}. \quad (13.6) \end{aligned}$$

#### Section 14.

We shall not discuss in detail the somewhat complicated integral of (13.5); one feature of it alone is relevant to our present purposes. When  $t_2$  is so great that  $s(t_2)=0$ , the events that determine one of the two cuts are not concerned in the other. Then there is no correlation between the cuts, and the chance that they will both occur should be the product of the chances that each of them should occur at a snap reading; that is to say, we should have

$$C(y_0, t_2) = C^2(y_0) \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.1)$$

and

$$C_1(y_0, t_2) = C(y_0), \quad . \quad . \quad . \quad . \quad . \quad . \quad (14.2)$$

where  $C(y_0)$  is given by (8.9), (13.6) reduces as it should to this form in this case; for all the  $\rho$ 's are then zero;  $\Delta=1$ ;

$$\Delta_{kl} = 0 (l \neq k), \quad = 1 (l = k) \quad . \quad . \quad . \quad . \quad . \quad (14.3)$$

This observation shows how greatly  $C(y_0, t_2)\tau_1$  differs from the chance—we shall call it  $C_0(y_0, t_2)\tau_1$ —that the *next* downward cut after the upward cut occurs in  $\tau_1$  about  $t+t_2$ , which is the chance we need if we are to calculate the mean duration of a hill. For, since the chance of a downward cut ( $t+t_2$ ) must increase indefinitely as  $t_2$  increases,  $C_0(y_0, t_2)$  must tend to zero as  $t_2$  increases. In order to find  $C_0$ , we should have to impose the further condition that  $y$  is greater than  $y_0$  in each infinitesimal interval between  $t$  and  $t+t_0$ ;  $C_0$  involves an infinitely multiple integral, which we have not succeeded in reducing and which therefore cannot be found even by the most laborious computation.

#### Section 15.

In these circumstances it is natural to seek some indirect method of finding  $C_0$ , and there may be a temptation to resolve it into the sums and/or products of simpler chances, in the manner so familiar in the elementary theory of probability. But this method involves the assumption that the events having the simpler chances are uncorrelated and their chances independent; the essence of our problem is that the events are correlated; these methods are therefore inapplicable. This fact is so important that it may be well to give an example of an argument that is *not* valid.

One method of deriving equations (1.6), (1.7) is to start by calculating

the chance  $W(n, T)$  that  $n$  of the events occur in the interval  $T^*$ . In order to find  $W(n, T)$ , we may use the proposition that the chance that  $(n+1)$  events occur in the interval  $T+dT$  is the sum of the chances (1) that  $n$  events occur in  $T$  and one in  $dT$ , and (2) that  $(n+1)$  events occur in  $T$  and none in  $dT$ . The chance that one event occurs in  $dT$  is  $\lambda dT$ ; the chance that two or more events occur in  $dT$  is of higher order in  $dT$  and may be neglected; so that the chance that none occur in  $dT$  is  $(1-\lambda dT)$ . Consequently we have

$$W(n+1, T+dT) = W(n, T)\lambda dT + W(n+1, T)(1-\lambda dT), \quad (15.1)$$

leading to a differential equation which, in conjunction with the "boundary" conditions  $W(0, 0)=1$ ,  $W(n, 0)=0$  if  $n \neq 0$ , enable us to find  $W(n, T)$ .

But the argument assumes that  $\lambda$  is independent of  $n$ , and that the chance of an event during  $dT$  is the same however many events have occurred in  $T$ . If there is correlation between the events, this assumption is untrue, and nothing is gained by writing (15.1), although it is still formally true if  $\lambda$  is understood to be a function of  $n$ .

#### Section 16.

Another alternative is to find some sufficient approximation to  $C_0$ . If  $t_m$  is the mean duration of a hill, which we want to find,  $C_0$  must clearly have a maximum in the neighbourhood of  $t_2=t_m$ . Again, when  $t_2$  is much less than  $t_m$ ,  $C_0$  must be nearly equal to  $C_1$ . For the only chance included in  $C_1$  and excluded from  $C_0$  is the chance that  $y$  reaches or cuts the level  $y_0$  two or more times in  $t_2$ ; this chance is clearly small if  $t_2$  is much less than  $t_m$ .

Further, we know that  $C_0$  must be zero for  $t_2$  much greater than  $t_m$ . Accordingly we know the ends of the curve relating  $C_0$  and  $t_2$ , and require only a knowledge of the curve in the neighbourhood of  $t_m$  in order to have a fair idea of the form of the curve.

Now when  $t_2$  is near  $t_m$ , it is improbable that there will be many cuts in  $t_2$ . We do not therefore need to impose the full condition that  $y > y_0$  in each infinitesimal interval within  $t_2$ ; it is sufficient to provide  $y > y_0$  at each of a few of those intervals. Further we know that the chance that  $y$  will attain  $y_0$  between an upward and a downward cut at  $y_0$  is less than the chance that it will attain  $y_0$  between two points at each of which  $y$  is merely equal to  $y_0$ . This latter chance is simpler to calculate than the former; its substitution for the former enables us to derive a lower limit to  $C_0$ . Since the purpose of this paper is to expound general principles rather than to attain definite results, these possibilities will not be examined further; but it appears that, with suitable forms of  $s(t)$ , they might lead to a sufficient approximation without impossibly laborious computation.

---

\* J. M. Whittaker, Proc. Camb. Phil. Soc. xxxiii. p. 451 (1937).



## Section 17.

Further properties of a trace may be obtained by considering its behaviour when certain conditions are imposed—which may be considered as “boundary conditions.” For example, the question may be asked: What is the chance of finding the trace between  $y$  and  $y+dy$  at time  $t$  if it is known that it passes through  $y_0$  at time  $t=0$ ? This may be answered quite easily.

The required chance is the chance of  $y_0$  at  $t=0$  and  $y$  at  $t$  occurring together divided by the chance of  $y_0$  occurring, *i. e.*

$$W(y, t) = \frac{\frac{1}{2\pi\sigma^2(1-\rho^2)^{1/2}} \exp \left[ -\frac{1}{2\sigma^2(1-\rho^2)} (y_0^2 - 2\rho y y_0 + y^2) \right]}{\frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{y_0^2}{2\sigma^2} \right]}. \quad (17.1)$$

$$= \frac{1}{\sigma(1-\rho^2)^{1/2}\sqrt{2\pi}} \exp \left[ -\frac{(y-\rho y_0)^2}{2\sigma^2(1-\rho^2)} \right], \quad \dots \dots \dots (17.2)$$

where

$$\sigma^2\rho = \lambda \int_0^\infty s(x)s(x+t) dx, \quad \dots \dots \dots (17.3)$$

and  $\rho$  is, of course, a function of  $t$ .

This is a Gaussian with mean  $\rho y_0$  and standard deviation  $\sigma(1-\rho^2)^{1/2}$ . When  $t=0$ ,  $\rho=1$ , and when  $t$  becomes large enough to make  $s(t)=0$ ,  $\rho \rightarrow 0$ , so that as  $t$  increases, the mean goes from  $y_0$  to 0, and the standard deviation goes from 0 to  $\sigma$ ; which is consistent with common sense.

It is perhaps worth while to work out a simple case.

$$\text{Let} \quad s(t) = e^{-\alpha t} - e^{-\beta t}, \quad \dots \dots \dots (17.4)$$

which satisfies the conditions  $s(0)=0$  and  $s(t) \rightarrow 0$  as  $t \rightarrow \infty$

$$\text{Then} \quad \sigma^2 = \frac{\lambda(\alpha-\beta)^2}{2\alpha\beta(\alpha+\beta)}, \quad \dots \dots \dots (17.5)$$

and

$$\rho = \frac{\beta}{\beta+\alpha} e^{-\alpha t} - \frac{\alpha}{\beta-\alpha} e^{-\beta t}. \quad \dots \dots \dots (17.6)$$

For  $\beta=2\alpha$ , fig. 2 *a* shows  $s(t)$ ; fig. 2 *b* shows  $\rho$  and fig. 2 *c* shows  $(1-\rho^2)^{1/2}$ . It is clear from fig. 2 *b* how the mean  $y_0\rho$  returns to the base line from  $y_0$ , and from fig. 2 *c* how the standard deviation  $\sigma(1-\rho^2)^{1/2}$  returns to  $\rho$  from zero as  $s(t) \rightarrow 0$ .

Now, from the definition of  $W(y)$  in Sections 4, 5 it follows that the proportion of the time spent by the trace between two levels  $y_1$  and  $y_2$  is given by

$$\int_{y_1}^{y_2} W(y) dy, \quad \dots \dots \dots (17.7)$$

since the chance of finding the trace between  $y_1$  and  $y_2$  must be proportional to the time spent by the trace between these levels.

Thus, after passing through  $y_0$ , the time, on the average, that the trace spends below the base line is

$$\int_{-\infty}^0 \frac{1}{\sigma(1-\rho^2)^{1/2}\sqrt{2\pi}} \exp \left[ -\frac{(y-\rho y_0)^2}{2\sigma^2(1-\rho^2)} \right] dy \quad . \quad . \quad . \quad (17.8)$$

$$= \frac{1}{2} \left[ 1 - \operatorname{erf} \left( \frac{\rho y_0}{\sigma \sqrt{2(1-\rho^2)}} \right) \right] \quad . \quad . \quad . \quad . \quad . \quad (17.9)$$

This goes from 0 when  $t=0$  to 0.5 as  $t \rightarrow \infty$ , as is again what would be expected. It is instructive to study the average behaviour of the trace after passing through a level  $y_0$  for an oscillatory form of  $s(t)$  such as

$$s(t) = e^{-\alpha t} \sin \omega t, \quad . \quad . \quad . \quad . \quad . \quad (17.10)$$

but we will not deal with this problem here.

Fig. 2, *a*, *b* and *c*.

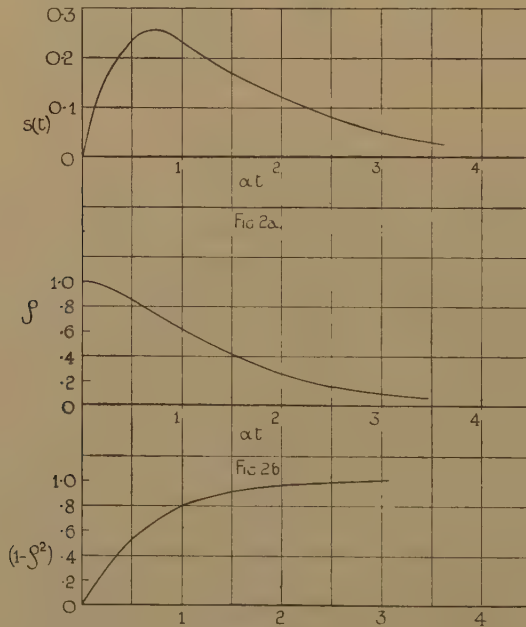


Fig 2c

Section 18.

The next simplest case to consider is the average behaviour of the trace after making an *upward* or a *downward* cut at  $y_0$ . The solution of this problem follows closely the analysis of Sections 9 and 10, correlating (1)  $y_1$  at  $t=0$  between  $-\infty$  and  $y_0$ , (2)  $\dot{y}_1 + ve$  at  $t=0$ , and (3)  $y$  at time  $t$ . The calculations are not difficult in this case. Other more difficult problems such as the average behaviour between two points a given time apart, through which the trace is known to pass, involve multiple integrals and are more difficult to deal with. These matters are being treated in another paper.

## Section 19.

In all the foregoing calculations it has been assumed that the deflections of the trace arise wholly from the events random in time (noise) and that there are no regular deflections (signal). If there are signals, probably the best way to represent the reaction between the signal and the noise is to make  $\lambda$  a function of  $t$  determined by the signal. This makes little formal difference to the calculations; for the equations, by means of which  $\lambda$  enters into our formulæ, remain true if  $\lambda$  is a function of  $t^*$ . But of course the mathematical difficulties are increased formidably.

## APPENDIX.

In Section 4 we arrived at the expression

$$W(y) = \sum_y \left[ \frac{(\lambda\tau)^{\sum n_r} e^{-\lambda T}}{II\{(n_r)!\}} \right], \quad \dots \quad (A.1)$$

where the summation  $\sum_y$  is taken over all sets of values  $n_r$  such that

$$\sum n_r s_r = y, \quad \dots \quad (A.2)$$

where  $\sum$  means summation over all values of  $r$  and where  $s_r$  is the value of  $s(t)$  during the  $r$ th interval according to (1.2). It should be remembered here that  $\sum n_r$  is the total number of events taking place in the time  $T$  defined in Section 1. We may assume without loss of generality that the  $s_r$  are all integral; for that merely involves choosing a sufficiently small unit in which to measure the  $s_r$ . Then, of course, the only possible values of  $y$  are integral. Let us write

$$e^{\mathbf{X}} = e^{\lambda\tau \sum s_r^2 r}, \quad \dots \quad (A.3)$$

Then the general term of  $e^{\mathbf{X}}$  is

$$\frac{(\lambda\tau)^{\sum n_r} x^{\sum s_r n_r}}{II\{(n_r)!\}} \quad \dots \quad (A.4)$$

and  $C(y)$ , the coefficient of  $x^y$  in  $e^{\mathbf{X}}$ , is

$$\sum_y \frac{(\lambda\tau)^{\sum n_r}}{II\{(n_r)!\}} \quad \dots \quad (A.5)$$

$W(y) = e^{-\lambda T} C(y)$ , so that the problem is to find  $C(y)$ . Now  $e^{\mathbf{X}}/x^{y+1}$  has a pole of order  $y+1$  at  $x=0$ ,  $C(y)$  is its residue at this pole. Hence, by the theory of residues,

$$C(y) = \frac{1}{2\pi j} \int_{\gamma} \exp\{\lambda\tau \sum s_r^2 r\} \frac{dx}{x^{y+1}}, \quad \dots \quad (A.6)$$

where the path of integration is any circle round the origin. If the variable is changed to  $\theta$ , the position angle of the radius vector, (A.6) gives

$$W(y) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \exp\{-\lambda T - jy\theta + \lambda\tau \sum e^{j\theta s_r}\} d\theta. \quad \dots \quad (A.7)$$

If the limit of  $s_r$  is so small that  $y$  may be treated as a continuous variable,

\* We owe this remark to a private communication from Prof. J. M. Whittaker.

and if  $T \rightarrow \infty$  (assumptions that involve no loss of generality), (A.7) becomes

$$W(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left\{ -jy\theta - \lambda \int_0^{\infty} (1 - e^{j\theta s(t)}) dt \right\} d\theta. \quad (A.8)$$

If the inner integral in (A.8) is expanded in powers of  $\theta$ , we have

$$W(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left\{ -jy\theta + j\theta \lambda \bar{s}_1 + \frac{\lambda(j\theta)^2}{2} \bar{s}_2 + \frac{\lambda(j\theta)^3}{3!} \bar{s}_3 + \frac{\lambda(j\theta)^4}{4!} \bar{s}_4 \dots \right\} d\theta, \quad (A.9)$$

where

$$\bar{s}_n = \int_0^{\infty} \{s(t)\}^n dt. \quad (A.10)$$

When  $\lambda$  is large the main contributions to the integral come from small values of  $\theta$ . If we neglect in (A.9) terms in powers of  $\theta$  higher than  $\theta^2$ , we have

$$W(y) \sim \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp \left\{ -jy\theta + j\theta \lambda \bar{s}_1 - \frac{\theta^2}{2} \lambda \bar{s}_2 \right\} d\theta. \quad (A.11)$$

This can easily be integrated, leading in the notation of (5.1) to

$$W(y) \sim G_1(y, \lambda \bar{s}_1, \sqrt{\lambda \bar{s}_2}), \quad (A.12)$$

i. e. a Gaussian distribution with mean  $\lambda \bar{s}_1$  and variance  $\lambda \bar{s}_2$ . This is the zero-th approximation given in (5.5). The neglected terms represent the higher approximations when  $\lambda \tau$  is not so large. These are easily obtained by expanding the part of the exponential containing terms of order higher than  $\theta^2$ . Thus

$$W(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left\{ 1 + \frac{\lambda(j\theta)^3}{3!} \bar{s}_3 + \frac{\lambda(j\theta)^4}{4!} \bar{s}_4 \dots \right\} \times \exp \left[ -jy\theta + \lambda j\theta \bar{s}_1 - \frac{(j\theta)^2}{2} \lambda \bar{s}_2 \right] d\theta. \quad (A.13)$$

Differentiating (A.11)  $n$  times with respect to  $y$ , we have

$$G_1^{(n)} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (-j\theta)^n \exp \left\{ -jy\theta + j\theta \lambda \bar{s}_1 - \frac{\theta^2}{2} \lambda \bar{s}_2 \right\} d\theta, \quad (A.14)$$

we obtain at once

$$\begin{aligned} W(y) = & G_1 - \frac{\lambda \bar{s}_3}{3!} G_1^{(3)} + \frac{\lambda \bar{s}_4}{4!} G_1^{(4)} - \frac{\lambda \bar{s}_5}{5!} G_1^{(5)} \\ & + \left\{ \frac{\lambda \bar{s}_6}{6!} + \frac{\lambda^2 \bar{s}_3^2}{2(3!)^2} \right\} G_1^{(6)} \\ & - \left\{ \frac{\lambda \bar{s}_7}{7!} + \frac{\lambda^2 \bar{s}_3 \bar{s}_4}{3! 4!} \right\} G_1^{(7)} \\ & \dots \dots \dots (A.15) \end{aligned}$$

### Reference.

- (1) N. R. Campbell and V. J. Francis, "A Theory of Value and Circuit Noise," Jour. I.E.E. xciii. (pt. iii.), p. 45 (1946).



XXXI. *The Structure of an Electromagnetic Field in the  
Neighbourhood of a Cusp of a Caustic.*

By T. PEARCEY \*.

[Received October 10, 1945.]

ABSTRACT.

It has been found possible to make a detailed mathematical and numerical study of the field structure at and near a line focus of a cylindrical electromagnetic wave train possessing any finite amount of cylindrical aberration of the first order. By a suitable choice of parameters all the possible wave-lengths and degrees of aberration can be expressed in terms of a single infinite integral of two independent parameters. Diagrams of the magnitude and phase of this integral are given, and only the final mathematical results are described.

---

INTRODUCTION.

UNDER conditions of intense refraction within inhomogeneous isotropic media, it is found that the ray systems set up by a transmitter become complicated by the existence of caustics, where focusing of adjacent rays occurs. These caustics appear as a set of surfaces, upon which there may also exist a number of lines along which the focusing effects are even more intense than upon any other part of the caustics. These lines are the cusps of the caustics.

In cases of intense refraction over wide regions, the field structure due to a transmitter can be calculated to a fairly close approximation from the principles of geometrical optics, except in the neighbourhood of caustics and cusps, where these principles are not valid. In these latter regions the field can be calculated only by wave theory. In general, however, any particular field of rays will consist of a number of independent superimposed systems, some of which are simple fields unaffected by the ray fields giving rise to the caustics. The entire field is then the vector sum of the effect of each of the independent pencils, so that the field due to the rays generating the caustic or cusps can be considered separately.

A further complication, occurring in the consideration of the field structure near caustics, is the effect of the curvature of the rays caused by

---

\* This paper is submitted by permission of the Chief Scientific Officer of the Ministry of Supply.

refraction of the radiation. If the change of refractive index of the medium within one wave-length is sufficiently small, the refraction will have little effect near the caustic. It is, however, possible to include the ray curvature effects by a suitable choice of co-ordinates.

The present study of the field structure in the neighbourhood of a cusp of a caustic was originally stimulated by a more general study of the propagation of electromagnetic waves through isotropic media, whose refractive index ( $\mu$ ) varies slowly within distances comparable with the wave-length, *i. e.*,  $\lambda/2\pi \cdot \text{grad} (\ln \mu) \ll 1$ , particular interest applying especially to the case of horizontal stratification, in which  $\mu$  is not monotonic but exhibits a minimum at some given level. Because of the latter interest, study has been restricted to the case of fields which possess cylindrical instead of axial symmetry.

Since it is possible to smooth out the principal ray curvature effects in slowly varying media, the problem of the field near the cusp reduces to that of studying, in detail, the field at a cusp produced by an advancing cylindrical wave-train which possesses any non-zero amount of cylindrical aberration of the first order. The study, therefore, immediately applies also to the case of the focus of a cylindrical lens possessing aberrations of the first order. The smoothing out of the ray curvatures can be carried out analytically by choosing a non-euclidean co-ordinate system whose metrical coefficients are all identical and are proportional to  $\mu$ .

In the author's knowledge there exists only one other attempt at a detailed study of the field near a cusp\*, and that applied to the case of axial symmetry with first order spherical aberration.

### 1. The Field Near a Caustic.

It is well known that the field structure near a simple caustic depends upon the curvature of the caustic in relation to the wave-length of the radiation, and that the nature of the variation of the field in a direction perpendicular to the caustic can be expressed in terms of the Airy Integral function, and that the increase of the field over that which would exist in free space is

$$\left. \begin{aligned} & \left( \frac{9x_0^3}{121\rho\lambda^2} \right)^{\frac{1}{3}} \cdot \text{Exp} \left( \frac{i\pi}{4} \right) \cdot \int_0^\infty \cos (t^3 + Zt) dt, \\ & Z = \left( \frac{24\pi^2}{\rho\lambda^2} \right)^{\frac{1}{3}} z, \end{aligned} \right\} \dots \dots (1)$$

where  $x_0$  is the distance of the wave front generating the caustic, and  $z$  is the perpendicular distance from the caustic in a direction leading away from the rays generating the caustic;  $\rho^{-1}$  is the difference of the curvatures of the caustic and the rays generating it, and  $\lambda$  is the wave-length. The properties of this function are well known and need no description; but it should be noticed that the gain increases with wave-length only as

---

\* Picht., *Ann. d. Phys.* 1936.

$\lambda^{-\frac{1}{2}}$  but increases as the curvature increases as  $\rho^{-\frac{1}{2}}$ . The variation of the structure depends upon  $Z$ . The maximum value attainable is

$$1.656 \operatorname{Exp} \left( \frac{i\pi}{4} \right) \cdot \frac{x_0}{(\lambda \rho^2)^{\frac{1}{2}}} \text{ at } Z=1.02.$$

## 2. The Field near a Cusp.

In order to facilitate calculations similar to those which can be carried out with the aid of the Airy Integral, a complete study of the field in the neighbourhood of a cusp has been made. It is found that the increase of the field near a cusp placed symmetrically about the origin of co-ordinates  $x$  and  $y$ , above the equivalent free space value for a plane wave, is given by the expression

$$\left( \frac{12x_0^2}{\pi\lambda\sigma} \right)^{\frac{1}{2}} \cdot \operatorname{Exp} \left( \frac{i\pi}{4} \right) \cdot \int_{-\infty}^{\infty} \operatorname{Exp} \{i(Yt + Xt^2 + t^4)\} dt, \quad (2)$$

where  $Y = \left( \frac{192\pi^3}{\sigma\lambda^3} \right)^{\frac{1}{2}} y, \quad X = \left( \frac{12\pi}{\sigma\lambda} \right)^{\frac{1}{2}} x,$

and  $\sigma$  is a parameter defining the shape of the caustic near the cusp, *i. e.*,

$$x^3 = -\frac{9\sigma}{8} y^2 \quad \text{or} \quad X^3 = -\frac{27}{8} Y^2. \quad (3)$$

The infinite integral of the expression (2) has been evaluated over a wide range of the variable  $X$  and  $Y$ , and diagrams of the modulus and phase of the function are illustrated by figs. 1 and 2 respectively. The dependence of the entire function (2) on wave-length is somewhat more rapid than for a simple caustic, and varies as  $\lambda^{-\frac{1}{2}}$ . It will be noticed that the dependence of  $Y$  upon  $\lambda$  is more important than that of  $X$ . Usually, however, the magnitude of  $\lambda^{\frac{1}{2}} \sigma^{\frac{1}{2}}$  is much larger than the magnitude of  $\lambda^{\frac{1}{2}} \sigma^{\frac{1}{2}}$ , so that the scale of the structure along  $x$  is much larger than the scale along  $y$ .

A principal focus is formed on the  $X$ -axis, and is not situated at the origin; being displaced from the origin by a number of wave-lengths equal to

$$-0.894 \left( \frac{\sigma}{\lambda} \right)^{\frac{1}{2}}. \quad (4)$$

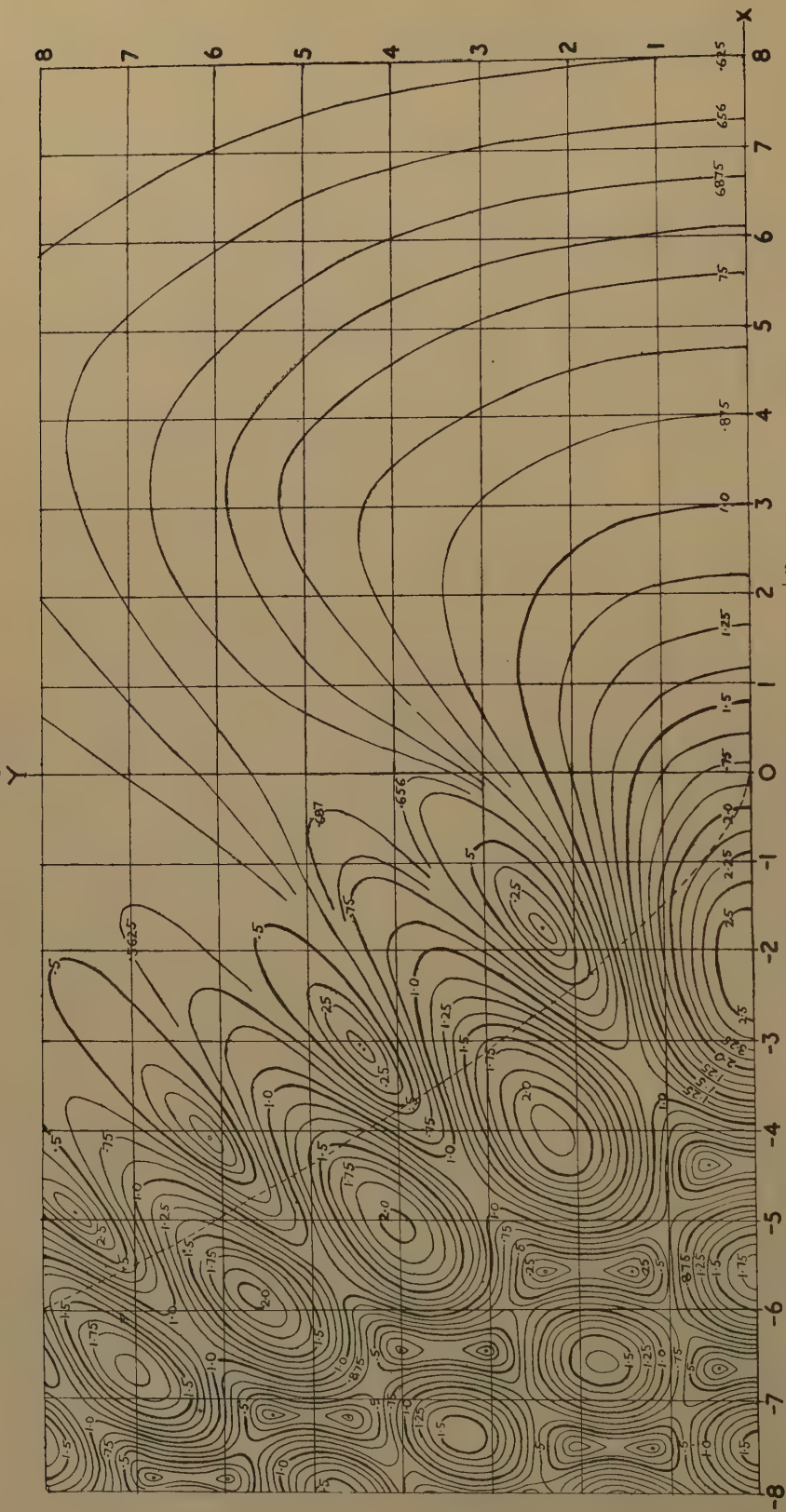
The maximum gain of the field over that due to a simple plane wave is

$$4.17 \left( \frac{x_0^2}{\lambda\sigma} \right)^{\frac{1}{2}}, \quad (5)$$

where  $x_0$  is the distance of the equivalent plane wave from the cusp.

To the right hand side of the caustic, the dotted line in the figures, there exists only one simple ray system, and the effects of the rays generating the caustics to the left decrease steadily with the distance to the right of the caustic. The hills and valleys on the left hand side are

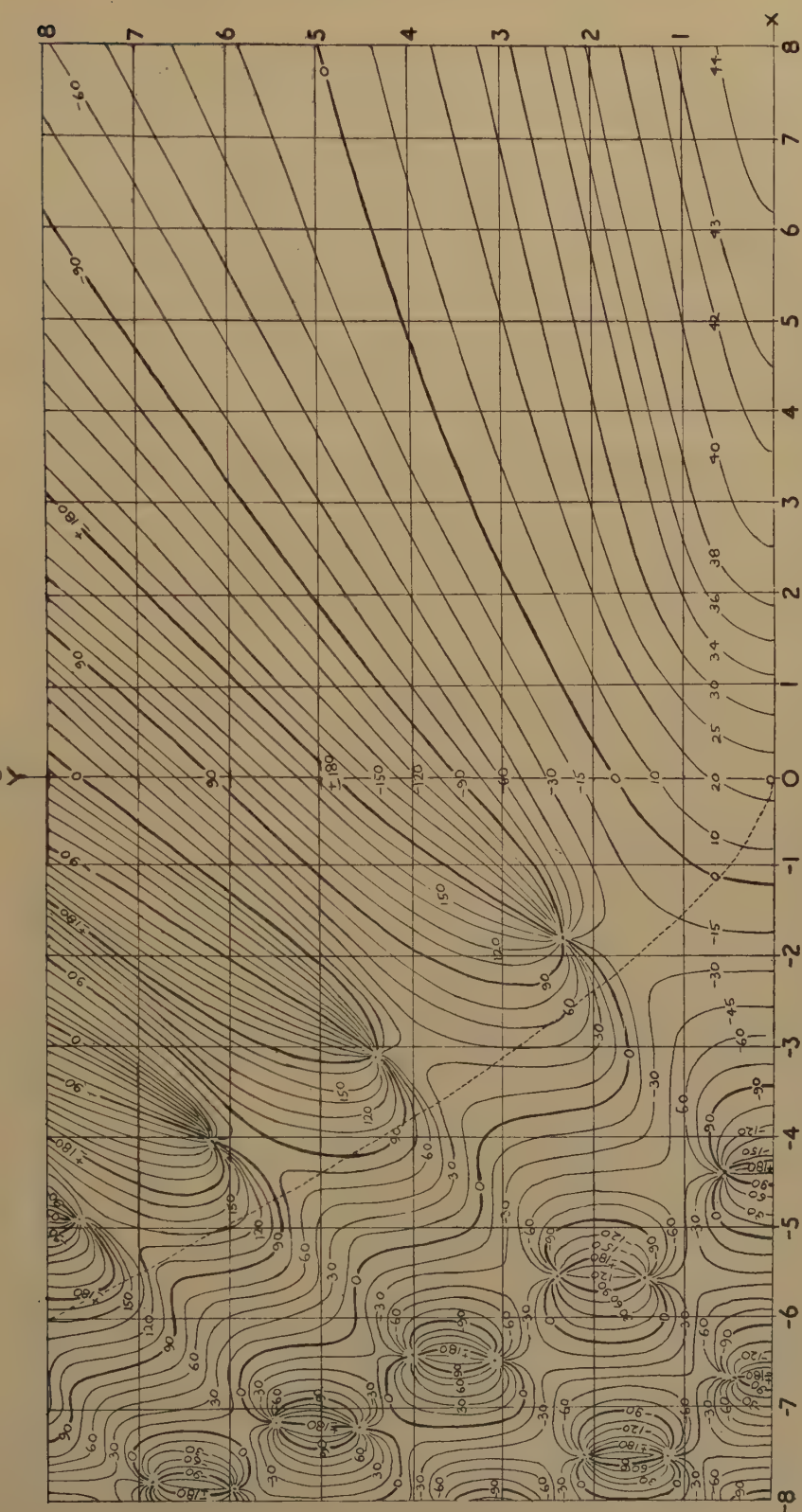
Fig. 1.



Contours of modulus of the cusp function  $I(X, Y) = \int_{-\infty}^{+\infty} \exp i(Yt + Xt^2 + t^4) dt$ .



Fig. 2.



Contours of phase of the cusp function (degrees)  $I(X, Y) = \int_{-\infty}^{+\infty} \exp i(Yt + Xt^2 + t^3) dt$ .

due to the interference of the individual fields of the three ray systems transversing that area. There are no interference effects far to the right, where the wave appears as though it were a cylindrical wave diverging from the origin. If we denote the integral part of (2) as  $I(X, Y)$ , then on the positive axis of  $X$  the function becomes

$$I(X, 0) \doteq \left(\frac{\pi}{X}\right)^{\frac{1}{2}} \text{Exp} \left(\frac{i\pi}{4}\right) \quad X \gg 0. \quad \dots \quad (6)$$

On the negative  $X$ -axis, to the left of the caustic, there are three ray systems, of which two, by symmetry, give identical effects. The effects due to the extra ray systems are superimposed upon the first set of rays, which converge toward the cusp and give rise to oscillations which increase in speed as  $|X|^2$  increases. They do *not* get small in comparison with average field, in fact,

$$I(X, 0) \doteq \left(\frac{\pi}{X}\right)^{\frac{1}{2}} \text{Exp} \left(\frac{i\pi}{4}\right) \cdot \left\{1 + i2^{\frac{1}{2}} \text{Exp} \left(\frac{iX^2}{4}\right)\right\} \quad X \ll 0. \quad \dots \quad (7)$$

The maximum at any point is  $2.41 \left(\frac{\pi}{X}\right)^{\frac{1}{2}}$ , and the minimum is  $0.41 \left(\frac{\pi}{X}\right)^{\frac{1}{2}}$ .

Along the  $Y$ -axis the effect of the caustics is to give rise to small oscillations, which rapidly disappear. The value of  $I(0, Y)$  varies finally as  $Y^{-\frac{1}{2}}$ , *i. e.*

$$I(0, Y) \doteq \left(\frac{2\pi^3}{27Y^2}\right)^{\frac{1}{2}} \left[ \text{Exp} \left\{i\left(\frac{\pi}{4} - \frac{3}{4} \frac{Y^{\frac{1}{2}}}{2^{\frac{1}{2}}}\right)\right\} + \text{Exp} \left\{-\frac{3^{\frac{1}{2}}}{8} \frac{Y^{\frac{1}{2}}}{2^{\frac{1}{2}}} + i\left(\frac{3}{8} \frac{Y^{\frac{1}{2}}}{2^{\frac{1}{2}}} - \frac{\pi}{3}\right)\right\} \right]. \quad \dots \quad (8)$$

The phase increases a little more rapidly than linearly.

Along the caustic, all the ray fields are important; the contribution of one pair decreases much more slowly than the contribution of the remaining one. The magnitude of the field finally decreases like  $Y^{-\frac{1}{2}}$ , whilst the oscillations die out more rapidly, like  $Y^{-\frac{3}{2}}$ . In the neighbourhood of the caustic far from the cusp, the structure must eventually depend primarily upon the field near a simple caustic, except that the field of the ray system from the opposite caustic will modify the total field. We, in fact, find \*,

$$I(X, Y) \doteq 24^{\frac{1}{2}} \frac{\pi}{|X|} \cdot \text{Exp} \left\{i\left(\frac{|X|}{2}\right)^{\frac{1}{2}}\right\} \cdot \left\{Y - 5\left(\frac{|X|}{2}\right)^{\frac{1}{2}}\right\} \cdot \text{Ai} \left[ \frac{\left(\frac{2|X|}{3}\right)^{\frac{1}{2}} - Y}{(8|X|)^{\frac{1}{2}}} \cdot z^{\frac{1}{2}} \right] \\ + \frac{0.1023 \dots}{|X|^{\frac{1}{2}}} \text{Exp} \left\{i\left(\frac{\pi}{4} - 2.331 \dots \frac{Y^{\frac{1}{2}}}{2^{\frac{1}{2}}}\right)\right\}, \quad \dots \quad (9)$$

so that eventually, at great distances along the caustic, the oscillations in its neighbourhood disappear, leaving only the caustic itself. This is

---

\* The function  $\text{Ai}(x)$  is the function tabulated by Dr. J. C. P. Miller.

as would be expected, as the rays from the opposite regions of the caustic suffer an extra divergence and diminution of effectiveness upon travelling to the opposite side.

Fig. 1 shows a number of sequences of zeros of  $I(X, Y)$ , ; these are distributed on either side of the lines "parallel" to the caustic, along which the first term of (9) is zero, and occur alternately on one side and then on the other as the relative phase of the two components of (9) vary by  $\pi$ . A single set of zeros occurs on the right-hand side of the caustic along the line upon which the magnitude of the two terms of (9) are equal.

The integral part of the expression (2), of which figs. 1 and 2 represent the modulus and phase respectively in terms of the cartesian co-ordinates  $X$  and  $Y$ , was computed in a series of steps. First, tables of the  $\frac{1}{4}$ -order Bessel functions were used for the evaluation of  $I(X, 0)$  and  $\left[ \frac{\partial^2 I(X, Y)}{\partial Y^2} \right]_{Y=0}$ . From a pair of differential equations satisfied by  $I(X, Y)$ , the function was evaluated along a number of lines lying parallel to the  $Y$ -axis by use of the Cambridge differential analyser. In the regions where the analyser was incapable of giving a sufficiently accurate solution the function was evaluated from its power series and its appropriate asymptotic expansions.

### 3. Use of the Diagrams.

If any given system of rays possesses a caustic and cusp system, and the structure of the field is required, the actual field intensity can be calculated with the aid of the figures in conjunction with equations (2). It is necessary only to know  $\lambda$  and to measure  $\sigma$  from the ray position. Besides illustrating the field structure in cases of intense non-uniform refraction, the figures show the character of the field at the focus of a cylindrical reflector possessing first order spherical aberrations, as well as the field near the focus of an optical system possessing cylindrical aberration.

### ACKNOWLEDGEMENT.

The author is greatly indebted to the staff of the Cambridge University differential analyser for collaboration in the computation involved in this work.

XXXII. *A Problem in Non-Steady Heat Conduction.*

By R. L. BROWN,  
Statistics and Mechanics Laboratory, the B.C.U.R.A.\*

[Received June 13, 1945.]

1. *Introduction.*

H. R. FEHLING † has discussed thermal insulation from the point of view of the non-steady state of heat transfer, and has shown how many important industrial insulation problems depend on an analysis of non-steady heat transfer. In particular he has discussed the case of a refractory wall, one face of which is raised continuously (and uniformly) in temperature whilst the other face is maintained at constant temperature. Another case of frequent occurrence is that of a refractory wall, to one side of which heat is supplied at a constant rate  $H$  (B.Th.U./ft.<sup>2</sup>hr.) whilst the other face is maintained at a constant temperature. This problem may be solved approximately by using the known solutions to the heat conduction equation when the temperatures on both faces of the wall are given. Here a general solution is given for the fraction  $h(t)$  of the final heat transfer at any time  $t$  after the heat supply  $H$  commences. So far as the author is aware this solution has not been previously reported.

2. *Statement of Problem.*

It is sufficient to take the case of an effectively infinite and flat wall with one face maintained at zero temperature. If  $K$  is the conductivity, the temperature  $\theta'$  of the heated face in the steady state is given by

$$H = K \frac{\theta'}{d}, \quad . . . . . (2.1)$$

where  $d$  is the wall thickness. The temperature  $\theta'$  is not reached instantaneously. Generally, if  $\theta$  is the temperature at time  $t$  and distance  $x$  from the heated face, we may find

$$\left( - \frac{\partial \theta}{\partial x} \right)_{x=d}$$

and thus evaluate  $h(t)$ .

The initial conditions are

$$t=0, \quad \theta=0 \text{ for all } x. \quad . . . . . (2.2)$$

---

\* Communicated by D. H. Bangham.

† H. R. Fehling, "Thermal Insulation," J. Inst. Fuel, p. 15 (Oct. 1944).



At time  $t$ , after the heat input  $H$  commences, we have

$$t > 0, \quad K \left( -\frac{\partial \theta}{\partial x} \right)_{x=0} = H, \quad \theta(d) = 0. \quad (2.3)$$

We require a solution of the heat conduction equation

$$\frac{\partial \theta}{\partial t} = D \cdot \frac{\partial^2 \theta}{\partial x^2} \quad (2.4)$$

in the form

$$h(t) = \frac{K}{H} \left( -\frac{\partial \theta}{\partial x} \right)_{x=d} \quad (2.5)$$

It will be assumed that the diffusivity  $D$  is independent of the temperature  $\theta$ .

### 3. The Temperature Gradient due to a Continuous Plane Source.

For an instantaneous plane source  $q$  at  $x=x'$ , Carslaw\* gives the solution of (2.4) as

$$\theta = \frac{q}{2\sqrt{\pi Dt}} e^{-(x-x')^2/4Dt}, \quad (3.1)$$

where the heat per unit area concentrated on the plane is  $(q \cdot K/D)$ . For a continuous plane source  $Q$  write  $q=Q \cdot dt'$  and  $(t-t')$  for  $t$ , and integrate equation (3.1) from 0 to  $t$ , obtaining

$$\theta = \frac{Q}{\sqrt{\pi D}} \int_{1/\sqrt{t}}^{\infty} \frac{1}{u^2} \cdot \exp \left\{ -\frac{(x-x')^2 u^2}{4D} \right\} du \quad (3.2)$$

by the transformation  $\frac{1}{\sqrt{t-t'}} = u$ .

Differentiating under the integral sign gives

$$\left( -\frac{\partial \theta}{\partial x} \right) = \frac{Q}{2D\sqrt{\pi D}} \int_{1/\sqrt{t}}^{\infty} (x-x') \exp \left\{ -\frac{(x-x')^2 u^2}{4D} \right\} du,$$

and writing

$$v = \frac{(x-x')u}{2\sqrt{D}}, \quad w = \frac{(x-x')}{2\sqrt{Dt}},$$

we obtain

$$\left( -\frac{\partial \theta}{\partial x} \right) = \frac{Q}{D\sqrt{\pi}} \int_w^{\infty} e^{-v^2} \cdot dv \quad (3.3)$$

Differentiation under the integral sign is clearly permissible, for the integrals in (3.2) and (3.3) both converge, whilst the latter converges uniformly for any value of  $x$ .

$$\text{Taking} \quad \operatorname{erf} w = \frac{2}{\sqrt{\pi}} \int_0^w e^{-v^2} \cdot dv, \quad (3.4)$$

we have finally

$$\left( -\frac{\partial \theta}{\partial x} \right) = \frac{Q}{2D} \cdot (1 - \operatorname{erf} w), \quad \text{where } w = \frac{(x-x')}{2\sqrt{Dt}}. \quad (3.5)$$

---

\* H. S. Carslaw, 'The Conduction of Heat in Solids,' p. 153.

4. *A System of Two Sources and Two Sinks.*

Consider a system symmetrical about  $x=0$ , consisting of

$$\left. \begin{array}{l} \text{sources} \quad +Q \text{ at } x' = 4nd \quad \text{and} \quad -4nd \\ \text{sinks} \quad -Q \text{ at } x' = -4n-2.d \text{ and } 4n-2.d \end{array} \right\} n \geq 1.$$

The temperature gradient at  $x=\lambda d$  is

$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=\lambda d} = \frac{Q}{2D} [-\operatorname{erf}(\overline{-4n-\lambda}.y) - \operatorname{erf}(\overline{4n+\lambda}.y) + \operatorname{erf}(\overline{4n-2+\lambda}.y) + \operatorname{erf}(\overline{-4n-2-\lambda}.y)], \quad (4.1)$$

where  $y = \frac{d}{2\sqrt{Dt}}$ . Now  $\operatorname{erf}(-w) = -\operatorname{erf}(w)$ , hence

$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=\lambda d} = \frac{Q}{2D} [\operatorname{erf}(\overline{4n-\lambda}.y) - \operatorname{erf}(\overline{4n+\lambda}.y) + \operatorname{erf}(\overline{4n-2+\lambda}.y) - \operatorname{erf}(\overline{4n-2-\lambda}.y)].$$

In particular,

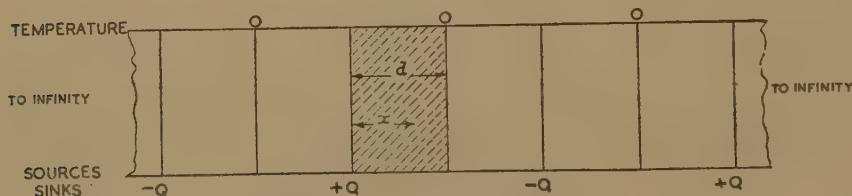
$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=0} = 0, \quad (n \geq 1), \quad \dots \quad (4.2)$$

$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=d} = \frac{Q}{2D} [-\operatorname{erf}(\overline{4n-3}.y) + 2\operatorname{erf}(\overline{4n-1}.y) - \operatorname{erf}(\overline{4n+1}.y)]. \quad (4.3)$$

5. *Solution of the Problem.*

Consider now a system of sources and sinks forming an infinite series (see fig. 1) of images reflected in  $x=0$  and  $x=d$ :

Fig. 1.



Infinite system of images.

$$+Q \text{ at } x = 0, \quad x = 4nd \text{ and } -4nd, \quad n \geq 1.$$

$$-Q \text{ at } x = -4n-2.d \text{ and } 4n-2.d, \quad n \geq 1.$$

The system is asymmetrical about  $x=d$ , so that we have  $\theta(d)=0$  for all time. The temperature gradients are, by (4.2) and (4.3),

$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=0} = \frac{Q}{2D} = \frac{H}{K},$$

thus satisfying (2.3)

$$\left( -\frac{\partial \theta}{\partial x} \right)_{x=d} = \frac{Q}{2D} \left[ \{1 - \operatorname{erf} y\} + \sum_{n=1}^{\infty} \{-\operatorname{erf}(\overline{4n-3}.y) + 2\operatorname{erf}(\overline{4n-1}.y) - \operatorname{erf}(\overline{4n+1}.y)\} \right],$$

which gives by (2.5)

$$\begin{aligned} h(t) = & (1 - \operatorname{erf} y) + (-\operatorname{erf} y + 2 \operatorname{erf} 3y - \operatorname{erf} 5y) \\ & + (-\operatorname{erf} 5y + 2 \operatorname{erf} 7y - \operatorname{erf} 9y) \\ & + (-\operatorname{erf} 9y + 2 \operatorname{erf} 11y - \operatorname{erf} 13y) + \text{etc.,} \quad (5.1) \end{aligned}$$

in which the remainder from the  $m$ th term onwards is

$$R_m = \sum_{n=m+1}^{\infty} p_n,$$

where  $p_n = \{-\operatorname{erf}(\overline{4n-3} \cdot y) + 2 \operatorname{erf} \overline{4n-1} \cdot y - \operatorname{erf} \overline{4n+1} \cdot y\}$ .

Now 
$$1 > \operatorname{erf} w > 1 - \frac{e^{-w^2}}{\sqrt{\pi} \cdot w},$$

and therefore

$$p_n < \frac{e^{-(4n-3)^2 y^2}}{\sqrt{\pi}(4n-3)y} + \frac{e^{-(4n+1)^2 y^2}}{\sqrt{\pi}(4n+1)^2 y^2} < \frac{2}{\sqrt{\pi}} \frac{e^{-16y^2(n-1)^2}}{4y(n-1)}.$$

Also 
$$\frac{e^{-16y^2 n^2}}{4yn} < \int_n^{n+1} \frac{e^{-16y^2 n^2}}{4yn} dn.$$

Hence 
$$\begin{aligned} R_m & < \int_m^{\infty} \frac{2}{\sqrt{\pi}} \frac{e^{-16y^2 n^2}}{4yn} dn = \int_{16y^2 m^2}^{\infty} \frac{1}{4y\sqrt{\pi}} \frac{e^{-u}}{u} du \\ & < \frac{1}{4y\sqrt{\pi}} e^{-16y^2 m^2} \log_e(16y^2 m^2) \end{aligned}$$

on integrating by parts. Hence, for  $y = 0.1$ ,  $R_m < 0.0001$ , for  $16m^2 y^2 > 8$ , i. e. for  $m > 7$ .

Thus the series in (5.1) converges rapidly, even for small values of  $y$ . Computation of the values of  $h(t)$  to 3 d.p. requires always fewer than five terms of the series; also for  $y > 0.6$  a very good approximation is given by

$$h(t) = (1 - 2 \operatorname{erf} y), \quad y > 0.6. \quad (5.2)$$

## 6. Application.

The variation of  $h(t)$  with  $(4Dt/d^2)$  is shown in fig. 2, from which the following table may be derived.

Time to reach various fractions of the final heat transfer.

$h \dots$	0.50	0.90	0.95	0.98	0.99
$4Dt/d^2 \dots$	1.5	4.1	5.2	6.7	8.0

It will be noted that to reach 99 per cent. of the final heat transfer takes about twice as long as to reach 90 per cent. Equilibrium is approached ( $h=0.99$ ) for times of the order of

$$t_1 = 2d^2/D. \quad (6.1)$$





XXXIII. *The Sensitivity of a Space-Charge Detector to the Presence of Positive Ions of Helium with Small Kinetic Energies.*

By D. M. MILLEST, Ph.D.\*

[Received August 21, 1945.]

ABSTRACT.

A method is described for investigating the effect of positive helium ions, with kinetic energies in the 25 to 125 electron-volt range, in augmenting the space-charge limited electron current from a hot cathode to a neighbouring anode. This increase is measured as a function of (a) the potential difference between anode and cathode, (b) the temperature of the cathode, (c) the kinetic energy of the positive ions, and (d) the pressure of helium in the apparatus. The lower limit of the positive ion currents which can be measured by means of the change which they produce in the space-charge limited electron current is also investigated.

---

*Introduction.*

IN the investigation of processes such as the ionization produced by collisions of neutral atoms, it may be difficult, when using an electrometer as the detecting instrument, to distinguish between the products of gaseous ionization and secondary electrons liberated from metal surfaces bounding the ionization chamber. The release of such electrons from the collecting electrode, in particular, may simulate the arrival of positive ions at that electrode. It is usually possible to carry out experiments which afford fairly conclusive evidence as to which process is taking place, but it would obviously be a great advantage to observe the onset of gaseous ionization by the detection of positive ions with a device which does not respond to the presence of electrons. Such considerations suggest the use of the "space-charge" detector developed by Kingdon<sup>(1)</sup> and by Hertz<sup>(2)</sup> for investigating the ionization produced by impacts of electrons with the atoms of a gas. It is well known that the action of this detector depends on the ability of the ions partially to neutralize the electron space-charge established in the neighbourhood of a hot cathode, the partial neutralization resulting in an increase in the space-charge limited electron current to a neighbouring anode. Although the space-charge detector has proved very suitable for revealing the onset of ionization in a gas subjected to electron bombardment or exposed to radiation<sup>(3)</sup>, its application to the study of atomic impacts has been less successful, and in some cases its use has led to results definitely contradictory to those obtained with an electrometer as the measuring instrument. Thus, Wayland<sup>(4)</sup> detected the ionization produced by argon

---

\* Communicated by Prof. Frank Horton, F.R.S.

atoms in neon, argon, krypton and xenon, using an electrometer, but obtained no measureable effect when the form of space-charge detector devised by Varney was employed. The ionization produced by impacts of helium atoms has been investigated with both instruments, and results obtained with a space-charge detector suggest that the minimum kinetic energy required by helium atoms in order to produce ionization in their own gas lies between 400 and 1,000 electron volts; for Varney<sup>(5)</sup> found no ionization to occur when the bombarding atoms had kinetic energies up to 400 electron volts, while Berry<sup>(6)</sup> did observe such ionization with atoms whose kinetic energies lay in the 1,000 to 8,000 electron volt range. The work of Rostagni<sup>(7)</sup> and of Horton and Millest<sup>(8)</sup>, on the other hand, gives a very different result. In these investigations the onset of ionization was observed with an electrometer, great care being taken to eliminate secondary effects, and in both cases ionization by atomic impact was observed with bombarding atoms of kinetic energy less than 100 electron volts, the minimum energy required being about 50 electron volts<sup>(8)</sup>.

Since the positive ions formed by atomic impacts probably acquire a considerable portion of the kinetic energy of the bombarding atoms, it seems desirable to investigate the sensitivity of a space-charge detector to positive ions of helium with kinetic energies in the 25 to 125 electron volt range. Information concerning the dependence of the sensitivity on the pressure of helium in the detector would also be useful, since if, in the study of atomic impacts, the collisions between swiftly moving and stationary atoms take place in the detector itself, pressures of the order of  $10^{-2}$  mm. of mercury or more may be employed.

Kingdon's investigations<sup>(9)</sup> indicated that the detector was sensitive to the presence of positive ions only when these were trapped within it, being made to oscillate to and fro across the region of dense electron space-charge, and prevented, by a suitable arrangement of potentials, from striking the electrodes bounding the detector. Kienzle<sup>(10)</sup> has shown, however, that when the kinetic energy of the positive ions is such that they are not confined to oscillatory paths, but can pass straight through the detector, a large increase in the space-charge limited electron current to the anode is obtained when the ion beam traverses the detector in such a direction that it passes through the potential minimum near the hot cathode. The detector used in the present investigation was designed with a view to obtaining the greatest sensitivity possible with the ion beams and pressures employed: it was therefore arranged, as described below, that as many as possible of the positive ions passed through the region of minimum potential.

#### *Description of Apparatus.*

A diagrammatic view of the experimental tube is shown in fig. 1. The tube was made of soda glass and the electrodes of gas free nickel-sheet and gauze, the nickel wires leading from them being spot welded to platinum for sealing through the glass. These glass to metal seals, shown for convenience in the diagram as lying in a straight line, were actually arranged at points on the circumference of a circle. The filaments

$F_1$  and  $F_2$  were of pure tungsten,  $F_1$  being the source of electrons for ionizing the helium in the apparatus and thus producing the positive helium ions, while  $F_2$ , which consisted of four fine wires connected in parallel, was the cathode of the space-charge detector. The apparatus was connected by a tube X, 1.5 cm. in diameter, to a Macleod gauge and pumping system, through a tap of wide bore, a bulb containing phosphorous pentoxide and a liquid oxygen trap to condense mercury vapour from the gauge. A second tube Y, sealed on to the opposite side of the

Fig. 1.

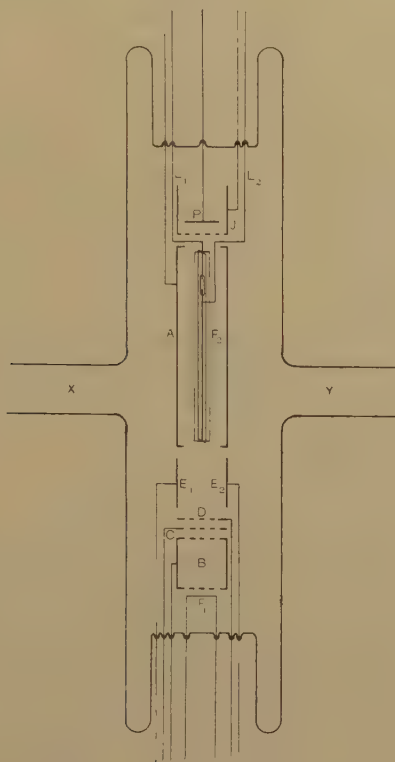


Diagram of experimental tube.

apparatus, served as an inlet for the helium, which was spectroscopically pure and entered the apparatus from a reservoir fitted with stopcocks and a tube containing activated charcoal, immersed in liquid oxygen.

Before admitting helium to the apparatus, the latter was baked for many hours with the filaments glowing and the high vacuum pumps running, in order to remove residual gas from glass and metal as completely as possible.

The filament  $F_1$  was zig-zagged in a horizontal plane and lay under the gauze covered base of the cylindrical electrode B. The gauze plate electrode C was mounted horizontally 2 mm. above B, and a similar plate electrode D was situated 2 mm. above C.  $E_1$  and  $E_2$  were two halves of a hollow cylinder 1 cm. long, which had been split in a vertical plane through the axis, and the adjacent edges separated by about 1 mm. in a



horizontal plane; their relative position was fixed by sealing a glass bead across each gap. About 2 mm. above the top of  $E_1$   $E_2$  was the lower end of the hollow cylindrical electrode A, 4 cm. long, which formed the anode of the space-charge detector. The cathode  $F_2$ , consisting of four tungsten wires mounted so that they formed equally spaced generators of a cylinder about 3 mm. in diameter, was coaxial with A. The nickel wire forming the axis of the framework across which the tungsten wires were stretched was cut at a point near its upper end, and the two parts sealed one into each end of a small glass bead. By welding the current leads  $L_1$  and  $L_2$  to the vertical wire on opposite sides of this bead, the four tungsten wires, in parallel, could be connected to the battery of accumulator cells which supplied the heating current. This design for the cathode was suggested by the following considerations:—if the detector consisted of a cylindrical anode with a single fine wire along its axis serving as cathode, the region of minimum potential between the electrodes, when the wire was glowing, would be situated very close to the surface of the wire. Hence only a small fraction of the incoming positive ions would enter that region where their presence was most effective in increasing the electron current to the anode, and those that did so would be very likely to collide with the cathode after traversing only a short distance. It was believed that, when the cathode shown in fig. 1 was made sufficiently hot, the region of dense space-charge would surround all four wires, thus increasing the fraction of the ion beam passing through the region of minimum potential, and also diminishing the probability that an ion travelling through this region would strike one of the wires. The cylindrical electrode J, the gauze base of which just cleared the horizontal portions of  $L_1$  and  $L_2$ , enclosed the plate electrode P.

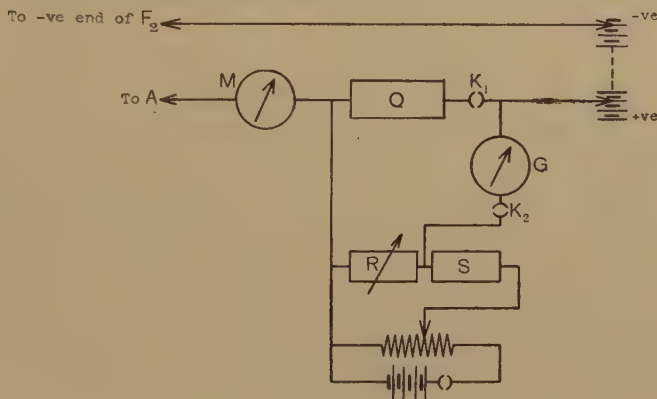
The positive ions, whose action on the electron space-charge established in the region of  $F_2$  was to be investigated, were produced by accelerating electrons from  $F_1$  through a potential difference of 40 volts into the field free region enclosed by electrode B, where they made collisions with neutral helium atoms. Some of the positive ions resulting from these collisions diffused through the gauze top of electrode B and were accelerated to C, the potential difference between B and C serving also to prevent electrons from  $F_1$  from reaching the latter electrode. The kinetic energy of the positive ions was controlled by the potential of electrode D below that of B, and the split cylinder  $E_1$   $E_2$  was maintained at the same potential as D. This cylinder was introduced in order that, by maintaining it at a potential slightly positive or negative to that of D, the beam of positive ions entering the space-charge detector could be made either to converge or diverge. It was thought that some increase in sensitivity might result from the use of a converging ion beam, in that a larger proportion of the ions would pass close to  $F_2$  and would therefore be more effective in neutralizing the potential minimum than if they travelled parallel to the axis of the tube, traversing the detector nearer to the anode A. Experiment showed, however, that such was not the case, and that the best results were obtained with  $E_1$  and  $E_2$  at the same potential as D. Probably a sufficient amount of convergence was obtained



by the potential difference between the anode and cathode itself. The negative end of the cathode  $F_2$  was made 1 volt positive with respect to  $E_1$ ,  $E_2$ , D and J, so that none of the electrons which it emitted when heated could escape from the detector, but were all collected at the anode A. The plate electrode P could be maintained at any convenient potential with respect to J, and was usually kept 6 volts negative with respect to this electrode, so that positive ions which had traversed the detector would be attracted to it.

The arrangements for measuring the electron current,  $i$ , between  $F_2$  and A, and the changes,  $\Delta i$ , produced in this current by the passage through the detector of a beam of positive ions, are shown in fig. 2.  $i$  was

Fig. 2.

Circuit for measuring  $i$  and  $\Delta i$ .

measured by the milliammeter M, the electrons then passing through the 500 ohm resistance Q. The potential difference between the ends of this resistance was balanced against that set up across the resistance R. The passage of a beam of positive ions through the detector increased the space-charge limited electron current passing through Q, and produced a deflection of the galvanometer G. This change,  $\Delta i$ , was measured by altering R until the deflection was again zero. Then

$$\Delta i = \frac{E}{Q} \times \frac{S(R_+ - R_0)}{(S + R_0)(S + R_+)},$$

where E is the potential difference across the two resistances R and S,  $R_0$  is the value of R for zero deflection of G when no positive ions enter the detector, and  $R_+$  the corresponding value when such ions are admitted.

The current  $i_+$  carried by the positive ions entering the detector was observed when  $F_2$  was cold and A made sufficiently negative with respect to  $F_2$  to collect all the incoming positive ions. The current was measured by G, the key,  $K_1$ , being open. While the effect of these ions on the electron space-charge round  $F_2$  was being investigated, the steadiness of the positive ion current was ensured by keeping constant the deflection of a galvanometer connected in the lead from either P

or C to the battery. Even at the lowest pressures investigated, the positive current collected at P in these circumstances was very much smaller than that measured by G as described above, and this difference increased at higher pressures. It appeared, therefore, that a large proportion of the ions entering the detector were collected at A,  $F_2$  or J. These would nevertheless contribute to the neutralization of the space-charge round  $F_2$  and ought to be taken into account; it was for this reason that the values of  $i_+$  were determined by driving the positive ions to A. Since the deflection of the galvanometer connected between C and the battery was unaltered by switching on the heating current through  $F_2$ , it was clear that no disturbances were produced in  $i_+$  by the electron emission from  $F_2$ , and that the measurements made with  $F_2$  cold were a reliable indication of the positive ion current entering the detector during the time when measurements of  $\Delta i$  were being carried out.

### *Experimental Procedure.*

If a beam of positive ions carrying a current  $i_+$  produces a change  $\Delta i$  in the electron current from  $F_2$  to A, we can define the amplification factor  $\mu$  of the detector as the ratio  $\Delta i/i_+$ . The primary object of this investigation was to determine how  $\mu$  depended on the following quantities:—

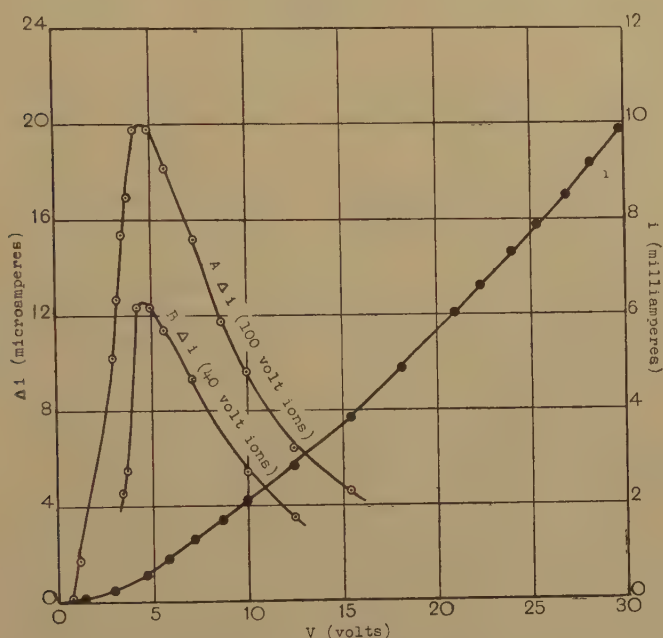
- (a) the potential difference between  $F_2$  and A,
- (b) the temperature of  $F_2$ ,
- (c) the kinetic energy of the positive ions,
- (d) the pressure of helium in the apparatus.

The dependence of  $\mu$  on the voltage  $V$  between A and the negative end of  $F_2$  was first investigated by fixing the potential difference accelerating the positive ions from B to D at some convenient value, and making A positive, step by step, with respect to  $F_2$ . The resistance  $R_0$  required to balance the drop in potential across Q when no positive ions entered the detector was determined for each value of  $V$  by reducing to zero the voltage  $V_F$  accelerating electrons from  $F_1$  to B, so that no positive ions were produced in the apparatus. On increasing  $V_F$  to 40 volts, so that positive ions were formed in B by electron impact, and entered A, the new resistance  $R_+$  which was required to restore balance was determined, and  $\Delta i$  calculated. That the change  $\Delta i$  was really due to the admission of positive ions to the detector, and not to any disturbing influence resulting from changing the potential of  $F_1$  with respect to B, was established by showing that  $\Delta i=0$  when this change was made with  $F_1$  cold, and also when the electron emission from  $F_1$  was increased to 5 milliamps while a high vacuum was maintained in the apparatus. Further, it was shown that the balance across Q was unaffected with  $F_1$  glowing, and with helium in the apparatus, so long as  $V_F$  was less than the ionization potential of helium. Thus the observed increase in  $i$  could be attributed entirely to the action of the positive ions of helium in the detector.

Fig. 3 shows  $i$  as a function of  $V$ , together with the corresponding values of  $\Delta i$  produced by positive ions of kinetic energies 40 and 100

electron volts. As  $V$  increases  $\Delta i$  rises sharply to a maximum, which it attains when  $V=4.6$  volts, and then falls off again. The value of  $V$  at which  $\Delta i$  reached its maximum was found to be independent of the kinetic energy of the positive ions over the range investigated, and to be independent also of the pressure in the apparatus. This independence of pressure is seen on comparing figs. 3 and 4; the former was obtained with the pressure of helium  $2.2 \times 10^{-2}$  mm. of mercury, and the latter with the pressure increased to  $10 \times 10^{-2}$  mm. of mercury; in each case  $V=4.6$  volts for a maximum  $\Delta i$ . The same curves, which were obtained

Fig. 3.



Showing  $i$  and  $\Delta i$  as a function of  $V$ .

Pressure =  $2.2 \times 10^{-2}$  mm. of mercury.

$i_+ = 0.0825$  microamperes, curve A.

0.0760 microamperes, curve B.

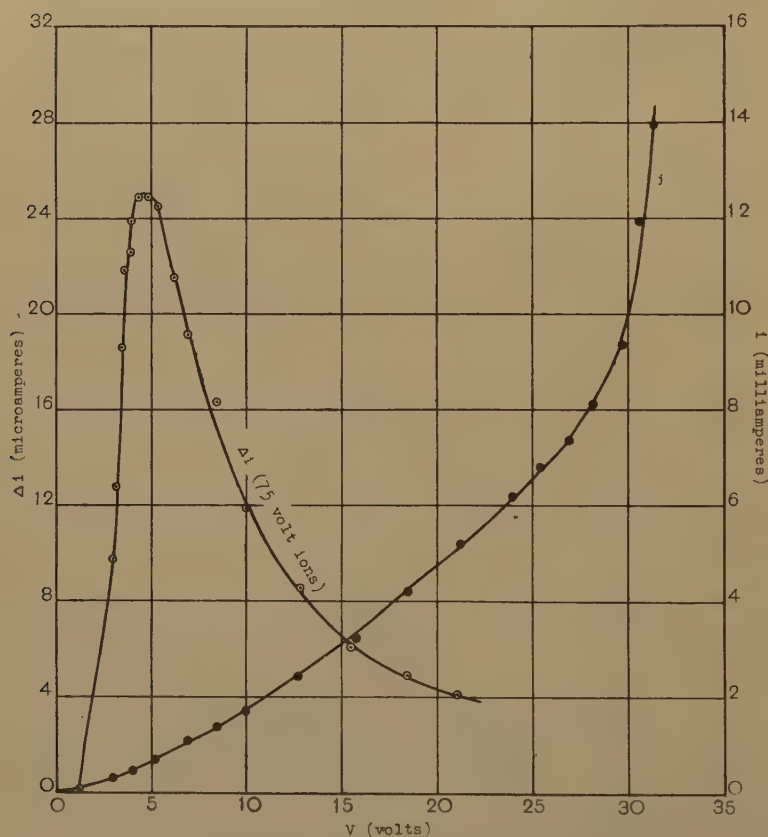
with 40, 75 and 100 electron volt ions, show that the optimum value of  $V$  does not depend on the kinetic energy of the positive ions.

The sudden increase of  $i$  when  $V$  exceeds about 27 volts is due to the ionization of helium by electrons from  $F_2$  (see fig. 4.) The difference between the voltage at which this increase begins and the actual ionization potential is due partly to the fact that  $V$  was measured with respect to the negative end of the cathode down which there was a potential drop of over 5 volts, and partly to the low pressure employed. When  $V$  is only a fraction of a volt greater than the ionization potential, the electrons will not acquire the amount of energy necessary for ionization until they have nearly reached A, and since the mean free path of the electrons is more than twice the distance from A to  $F_2$ , the probability of their

making an ionizing collision in the remainder of their journey is very small. Both these factors tend to make the number of ions produced when  $V$  is only just in excess of the ionization potential of helium extremely small.

As the temperature of  $F_2$  was diminished, the value of  $\Delta i$  corresponding to a given value of  $V$  diminished also, and the peak in the curve showing  $\Delta i$  as a function of  $V$  occurred at lower voltages. The curves in figs. 3 and 4, showing a peak when  $V=4.6$  volts, were obtained with a heating

Fig. 4.



Showing  $i$  and  $\Delta i$  as a function of  $V$ .  
 Pressure  $10 \times 10^{-2}$  mm. of mercury.  
 $i_+ = 0.275$  microamperes.

current of 1.14 amperes through each of the four filaments comprising  $F_2$ : on reducing this current to 0.9 ampere, the peak was obtained when  $V=4.2$  volts. Smaller heating currents than this were not used, because  $\Delta i$  became too small to measure accurately.

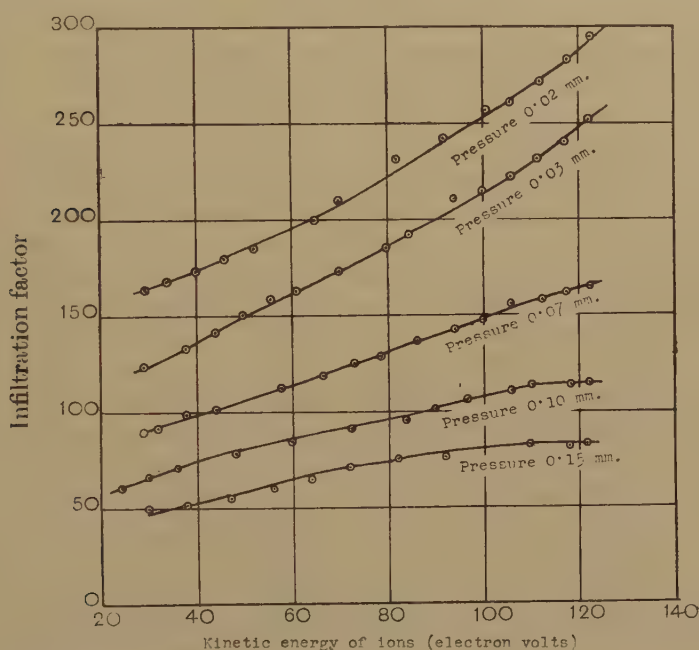
The amplification factor of the detector, for a beam of helium ions of given kinetic energy, and for a fixed value of the pressure, depends therefore on both the anode voltage and the temperature of the cathode.



The values stated in the present paper are those obtained with the cathode at the maximum temperature which could be used safely for a long period of time, and with the anode voltage at the optimum value for this temperature (4.6 volts).

The variation of the amplification factor with the kinetic energy of the ions was next investigated,  $V$  being fixed at 4.6 volts, and the potential difference between C and D altered step by step. The current  $i_+$  carried by the positive ions was measured on the galvanometer G for each value of the accelerating voltage before and after the set of observations of  $\Delta i$ . The steadiness of  $i_+$  during these measurements of  $\Delta i$  was indicated by a galvanometer connected between electrode C and the battery.  $\Delta i$  was

Fig. 5.



Showing the amplification factor as a function of the kinetic energy of the positive ions.

determined as in the previous observations, the value of  $R$  for balance across  $Q$  with  $V_F=0$  being checked after each individual determination. In fig. 5 the relation between the amplification factor  $\mu$  and the kinetic energy of the positive ions entering the detector is shown for five different values of the pressure of helium. At each of the pressures investigated  $\mu$  is an increasing function of the kinetic energy of the ions, and for a given value of the energy  $\mu$  increases as the pressure is diminished.

#### Discussion of Results.

Kienzle<sup>(11)</sup> has shown that a beam of positive ions with kinetic energy in the 100 to 1250 electron volt range has a marked effect on the space-

charge limited electron current from a hot cathode only when the ions pass very close to the cathode, and thus traverse the region of minimum potential resulting from the negative space-charge round it. He therefore concludes that positive ions are effective in increasing the electron current when, by partially neutralizing the negative space-charge in the region of the potential minimum, they lower the barrier which this minimum presents to electrons travelling from the cathode towards the anode. This conclusion is supported by the results of an investigation by Varney<sup>(12)</sup>, who showed that when the ion beam passed through the detector parallel to the length of the glowing filament, and at a distance of a few millimetres from it, only a very small effect was obtained with a high vacuum in the apparatus, but that, on admitting gas so that some of the positive ions were made to pass closer to the filament as a result of collisions with stationary atoms, and to cross the region of minimum potential, a much greater increase in the electron current to the anode resulted. On the above view of the action of the positive ions, it is possible to account for the existence of an optimum anode voltage, at which a beam of positive ions of given intensity produces a maximum increase in the electron current, and for the shift of this optimum value to higher anode voltages as the temperature of the cathode is increased. Consider a beam of positive ions with given velocity, traversing the detector in a fixed direction, the pressure of helium in the detector remaining constant. It is reasonable to suppose that the amount by which the potential barrier round the cathode is lowered by the passage of this beam depends only on the intensity of the beam, and not on the actual height of the barrier. For a given temperature of the cathode, the height of the potential barrier decreases as the anode voltage  $V$  is increased; hence, on the above supposition, it is to be expected that the diminution in height produced by the positive ions would be independent of  $V$  until the latter became so large that the barrier could be completely destroyed by an ion beam of smaller intensity than that actually employed. The electrons which, in the absence of the ion beam, manage to cross the barrier and reach the anode, thus giving the observed current  $i$ , are those whose kinetic energy, after escaping from the surface of the cathode, is greater than the height of the barrier expressed in energy units. The admission of a beam of positive ions of constant intensity will therefore result in the greatest increase,  $\Delta i$ , in the anode current when the height of the barrier is such that a given small diminution in height results in the maximum possible increase in the number of electrons able to overcome it. This will occur when the height is of the same order of magnitude as the most probable energy of emission of electrons from the cathode. Since the height of the barrier depends on  $V$ , the presence of a maximum in the curve showing  $\Delta i$  as a function of  $V$  is accounted for. When the temperature of the cathode is diminished, the height of the barrier for a given anode voltage is reduced also, and on this account the maximum in the  $\Delta i$ — $V$  curve occurs at a lower value of  $V$ . The shift is small, because at the lower temperature the most probable energy of emission of electrons is diminished also, and this tends to shift the maximum in

the opposite direction. The decrease in height of the maximum with diminution of temperature is to be expected, since at lower temperatures the cathode is emitting fewer electrons. In connection with this explanation of the presence of an optimum anode voltage, it is interesting to notice that, in those experiments in which he used a tungsten filament, Kienzle obtained a maximum in his  $\Delta i$ - $V$  curves when  $V=5$  volts, which is not very different from the optimum value obtained in the present investigation. The temperature of the cathode in Kienzle's detector was about the same as that of  $F_2$  in the present investigation, but the other conditions, such as the nature of the positive ions, their velocity, and their direction of motion with respect to the length of the cathode, as well as the design of the cathode itself, were very different. Thus the similarity of the values obtained for the optimum anode voltage in the two cases supports the suggestion that the relation between the most probable energy of emission of electrons, and the height of the potential barrier which they have to overcome once they have left the cathode, in order to reach the anode, is the most important factor governing the effect which the admission of a beam of positive ions will have on the space-charge limited electron current to the anode.

The increase in the amplification factor with the velocity of the positive ions, the anode voltage remaining fixed at its optimum value, is probably due to the deflection of the ion beam towards the cathode in the electric field established between this electrode and the anode of the detector. When the kinetic energy of the positive ions is small, they are deflected sharply towards the cathode on entering the detector and pass rapidly out of the region of dense space-charge, either being captured on the cathode or else passing across it to strike the anode on the opposite side. The greater the kinetic energy of the ions, the smaller is the inclination of their path towards the cathode, and the longer is its length in the region of minimum potential between the electrodes. At the higher pressures used, the proportion of slow ions entering the detector, for a given value of the accelerating voltage, is large on account of collisions made with stationary atoms; hence it is to be expected that, when the voltage accelerating the positive ions is fixed, the amplification factor should diminish as the pressure is increased.

*The Determination of the Minimum Positive Ion Current which can be Measured with the Space-charge Detector.*

Theoretically, the minimum positive ion current which can be detected is restricted only by the maximum resistance which it is convenient to use for  $Q$ , and the sensitivity of the galvanometer  $G$ ; but the method is limited in practice by the difficulty in eliminating random fluctuations in the electron emission from  $F_2$ . For example, if the kinetic energy of the ions is 30 volts, and the pressure of helium in the detector is  $2 \times 10^{-2}$  mm. of mercury, the amplification factor for  $V=4.6$  volts is 164 (see fig. 5). Thus if  $i_+ = 10^{-10}$  ampere,  $\Delta i$  will be  $1.64 \times 10^{-8}$  ampere, and, if reliable measurements are to be obtained, the random fluctuations



in  $i$  should not exceed one-tenth of this value—i. e. they should not be greater than about  $1.6 \times 10^{-9}$  ampere.  $i$  itself was  $7 \times 10^{-4}$  ampere, and hence must be kept constant to one part in 400,000. It was just possible to detect positive ion currents of the order of  $10^{-10}$  ampere by adjusting  $R$  so that the deflection of  $G$  was small when  $V_F$  was zero, then switching  $V_F$  over to 40 volts, observing the change in deflection of  $G$ , and checking the original setting by making  $V_F$  zero again. In this way it was necessary for  $i$  to remain constant to within the specified limits for periods of about 30 seconds only. This extra adjustment makes the method rather laborious to use, and even so the sensitivity is greatly inferior to that afforded by an electrometer. To detect currents of the order of  $10^{-14}$  ampere, which can be measured on a sensitive electrometer, it would be necessary to reduce random fluctuations in  $i$  to less than one part in  $10^9$ —a degree of constancy which it was found not possible to achieve even for a few seconds.

Kienzle's work has shown the form of space-charge detector hitherto used for observing the onset of ionization by atomic impact to be insensitive because the ions produced did not pass through the region of minimum potential between the hot cathode and the anode. The results of the present investigation suggest that, even when conditions are arranged so that an ion beam, by passing very close to the cathode for a considerable distance, has the greatest possible effect in augmenting the space-charge limited electron current to the anode, the space-charge detector is still less suitable than an electrometer for detecting small numbers of positive ions with the kinetic energies and at the pressures likely to be encountered in the investigation of the onset of ionization by atomic impact.

#### Acknowledgment.

In conclusion, I should like to express my thanks to Professor Horton for his advice and encouragement throughout the course of this investigation.

#### References.

- (1) Kingdon, *Phys. Rev.* xxi. p. 408 (1923).
- (2) Hertz, *Z. Phys.* xviii. p. 307 (1923).
- (3) Foote and Mohler, *Phys. Rev.* xxvi. p. 195 (1925).
- (4) Wayland, *Phys. Rev.* lii. p. 31 (1937).
- (5) Varney, *Phys. Rev.* l. p. 159 (1936).
- (6) Berry, *Phys. Rev.* lxii. p. 378 (1942).
- (7) Rostagni, *Nuovo Cim.* xi. p. 621 (1934).
- (8) Horton and Millest, *Proc. Roy. Soc. (A)* clxxxv. p. 381 (1946).
- (9) Kingdon, *loc. cit.*
- (10) *Ann. der Phys.* 5, xxx. p. 403 (1937).
- (11) Kienzle, *loc. cit.*
- (12) Varney, *Phys. Rev.* liii. p. 732 (1938).

Royal Holloway College,  
University of London,  
Englefield Green,  
Surrey.



XXXIV. *An Experimental Study of the Quantum Nature of X-rays.*

By M. P. GIVENS,

Physics Department, The Pennsylvania State College,  
State College, Pa.\*.

[Received April 29, 1946.]

*Abstract.*

A study has been made of the X-rays symmetrically diffracted by the (130) and ( $\bar{1}\bar{3}0$ ) planes of NaCl. The object of the study was to determine whether or not X-rays are diffracted in both directions simultaneously. Two Geiger Müller counters were used to detect the X-rays. A coincidence circuit was connected to the two counters. Over 20 million quanta were counted by each counter, yet only 12 thousand apparent coincidences were observed. A detailed study of the data reveals that this small number of coincidences is best interpreted by assuming no coherence between the two diffracted X-ray beams and attributing the coincidence counts to the chance arrival of two quanta, one at each counter, separated in time by less than the resolving time of the coincidence circuit. This lack of coincidence is in accord with the quantum picture of radiation.

1. *Introduction.*

THE purpose of this work is to obtain direct experimental evidence for or against the quantum nature of X-rays. The general plan of the experiment is shown in fig. 1. X-rays are produced at X, and after passing through pin holes in lead sheets at A and B fall upon a rock salt crystal. The crystal is so oriented that the X-rays arrive along a normal to the (100) planes. The X-rays are diffracted to the left and right by symmetrical planes (130 and  $\bar{1}\bar{3}0$ ), and after passing through the slits S and S' reach one of the Geiger-Müller counters  $GM_L$  or  $GM_R$ . Lead shielding, not shown in the figure, provides the necessary protection.

The quantum picture provides a certain probability that any given X-ray quantum may be diffracted either to the left or to the right (but not in both directions simultaneously), and that probability is independent of what any other X-ray quantum may do. On the basis of the quantum picture, it would be predicted that the arrival of X-ray quanta at  $GM_L$  and  $GM_R$  should be incoherent. If the two counters  $GM_L$  and  $GM_R$  are connected to a coincidence circuit, the coincident counts should not

\* Communicated by Sir Lawrence Bragg, O.B.E., F.R.S.

be in excess of those resulting from chance coincidences or "near coincidences" which are counted because of the finite resolving time of the circuit. On the other hand, if the number of coincidences be too great to be explained in this way, the experiment would be an indication that the X-rays were diffracted in both directions (left and right) simultaneously. This would contradict the quantum theory. Such a result would require that the number of coincidences should equal the number of counts at either counter, subject to a correction for the fact that the counters are not 100 per cent. efficient.

Experimentally it is found that out of 20 million quanta recorded at each of the two counters only 12,000 apparent coincidences were detected. It will therefore be necessary to examine, from as many points of view as

Fig. 1.

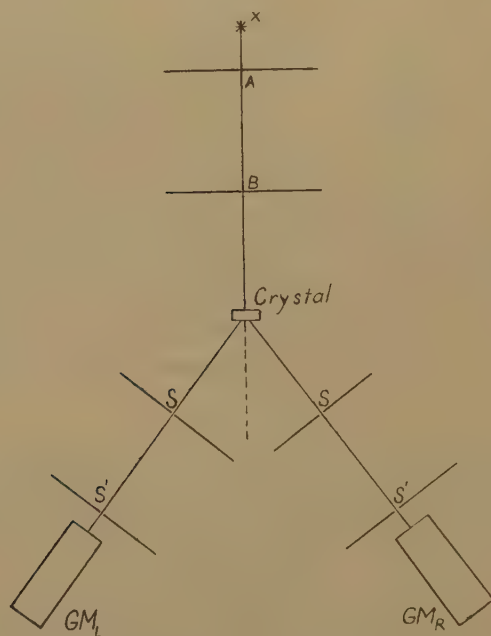


Diagram of apparatus.

may suggest themselves, the apparatus and the experimental data in order to see whether this small number of apparent coincidences can best be accounted for on the basis of coherent or incoherent diffraction. It is with such a discussion that the rest of the paper has to do.

## 2. *Mathematical Development.*

Two extreme cases will be considered : (A) the case in which complete lack of correlation is assumed, and (B) the case in which the X-rays are always diffracted so as to give a coincidence.

(A) Poisson's probability law\* may be applied to the case in which complete lack of correlation between the action of the two counters is assumed. This probability law is applicable when the probability of an event is small. It may be written in the form

$$p(n) = \frac{a^n e^{-a}}{n!}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where  $p(n)$  is the probability of  $n$  events in a given time and  $a$  is the average number of events in that time.

First let  $n=1$  and  $a=Nt$ , where  $N$  is the average number of events per unit time and  $t$  is some length of time which is arbitrary except that  $Nt$  must be much less than unity. Then

$$p(t) = Nte^{-Nt}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which gives the probability of observing one event in a specified short time interval  $t$ . If this is identified with the events at the left hand GM counter, the corresponding equation

$$p'(1) = N'te^{-N't} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

may be applied to the events at the right hand counter, where the primes indicate the right hand system. No prime is on  $t$ , since both counters are to be observed during the same short time interval.

The probability of observing one event at each counter is given by the product

$$p(1)p'(1) = NN't^2e^{-(N+N')t} \quad (4)$$

If  $t$  is now identified with the resolving time of the coincidence circuit, this equation gives the probability of a chance coincidence count in specified short time  $t$ . The probability of a chance coincidence count in unit time is  $C$ , where

$$C = \frac{2}{t} p(1) p'(1) = 2NN'te^{-(N+N')t} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The factor 2 enters, since the probability of two events happening within a time  $t$  of each other is twice as great as the probability that both will happen in a specified time interval  $t$ .

For the calculation of the experimental results the exponential factor was omitted. The error introduced by this simplification was approximately 0.2 per cent. Equation (5) then reduces to

$$t = \frac{C}{2NN'}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where  $C$  may be regarded as the average number of coincidence counts per unit time (sec.), the average to be taken over a long time;  $N$  and  $N'$

\* See for example, M. C. Holmes, 'An Outline of Probability and its Uses, p. 36. (Burgess Pub. Co. (1936); or H. C. Plummer, 'Probability and Frequency,' p. 182. (Macmillan and Co., 1940.)

are the average number of events at the left and right hand counters per unit time (sec.),  $t$  is the resolving time (sec.) which must be assigned to the coincidence circuit in order to attribute all the coincidence counts to chance.

If the value of  $t$  is constant for large variations in the counting rates  $N$  and  $N'$ , this fact will indicate that  $t$  is a characteristic of the circuit and that the coincidences are only chance coincidences. If a value of  $t$  for the circuit is obtained without the use of X-rays, it can be compared with the value of  $t$  from the X-ray data, and if the two are substantially the same this fact, too, will indicate only chance coincidences among the X-ray quanta. The data, which will be presented later, conform with these predictions and thus support the assumption of complete lack of correlation between the two counters.

(B) Now consider the other extreme case and assume that the X-rays are always diffracted in such a manner as to give a coincidence. Such a coincidence will not always be observed because the GM counters are not 100 per cent. efficient. Let  $E$  and  $E'$  represent the efficiencies of the left and right hand counters respectively, and interpret this as meaning that there is a random probability  $E$  (or  $E'$ ) that the counter will respond to a given quantum, independent of the past history of the counter or the quantum. Under these assumptions, whenever a quantum arrives at the left counter, there is a probability  $E$  that it will be counted, and at the same time there is a quantum at the right counter with a probability  $E'$  that it will be counted. The probability of a coincidence count is then  $EE'$ . If  $M$  represents the average number of coincidences arriving at each counter per unit time (sec.), then the average number of coincidence counts per unit time will be

$$C = EE'M. \quad \dots \dots \dots (7)$$

The number of counts at the left counter is  $N = EM$  and the number at the right counter is  $N' = E'M$ . Using these equations to eliminate  $M$  from equation (7), the following result is obtained :

$$E = \frac{C}{N'} \quad \text{and} \quad E' = \frac{C}{N}. \quad \dots \dots \dots (8)$$

The two counters were of identical construction and, since  $N$  and  $N'$  were always of approximately the same magnitude, the study was simplified by using the average value of the efficiency,

$$\bar{E} = \frac{1}{2}(E + E') = \frac{1}{2}C \frac{N + N'}{NN'} \dots \dots \dots (9)$$

If the efficiency of the counters, calculated from equation (9), has a reasonable value and is reasonably constant when  $N$  and  $N'$  are changed, this fact may be taken as an indication of true coincidences and a contradiction of the quantum picture. The data given below do not constitute such a contradiction of the quantum picture.



### 3. Apparatus.

The general plan of the apparatus has already been given in fig. 1. The two GM counters were connected to circuits which had previously been developed in this laboratory \*. Since only minor changes have been made in the circuits they will not be described here. A coincidence circuit of the Rossi type †, using type 6SF5 tubes, was connected to the two GM tubes through cathode follower circuits ‡.

The counter tubes, which are described elsewhere §, contain krypton and ether. The two sides of the system were made as nearly identical as possible. The X-ray tube was of the Coolidge type and contained a molybdenum target. It was operated from a half wave rectifier at 37 KV<sub>max</sub>.

Pin holes rather than slits are necessary at A and B, since slits might permit the X-rays reaching GM<sub>L</sub> and GM<sub>R</sub> to have started from widely separated points in the X-ray tube, thus preventing the detection of any coherence which might exist between the two diffracted beams. The pin holes were 1 m. in diameter and were 15 cm. apart. These pin holes provided an X-ray beam of finite angular spread (more than 0.1° half angle). This produced a spread in Bragg angle which made the position of the crystal less critical.

At S and S' slits were used, since pin holes were considered unnecessary in these locations. These slits were opened very wide (3 mm.) so as to allow the entire diffracted beam to enter the counter.

Optical methods (using visible light) were used to adjust the spectrometer so that the incident beam (through A and B) and the lines of sight of the two GM tubes would intersect over the centre of the spectrometer table and be coplanar (horizontal plane). The crystal was then placed at the intersection of these lines and was so oriented that the two diffracted beams would be deviated equal amounts (approximately 37°). This position of the crystal is called the "zero position." The crystal was mounted on a goniometer which enabled the crystal to be rotated about three mutually perpendicular axes, one of which was parallel to the incident beam and one was vertical. The orientation of the crystal about each of these axes could be determined by a vernier scale which could be read without interpolation to 0.1°. The two spectrometer arms were also equipped with vernier scales which could be read without interpolation to 0.1°.

### 4. Experimental Procedure and Results.

The coincidence circuits were tested by applying simultaneous pulses to the grids of the two first tubes. The pulses were provided from a

---

\* Davey, Smith and Harding, *Rev. Sci. Inst.* xv. p. 37 (1944).

† B. Rossi, *Nature*, cxxv. p. 636 (1930).

‡ See any recent book on high frequency circuits, *e.g.* Brainerd, Kohler, Reich, Woodruff, 'Ultra-High-Frequency Techniques,' p. 220. (D. Van Nostrand Co., 1942.)

§ Davey, *et. al.*, *loc. cit.*

battery through a resistance-condenser network which applied pulses of size and time duration approximately the same as the pulses provided by the GM tubes. The circuit never failed to respond to coincident pulses at the two tubes and never responded to a pulse applied to a single tube. From its remarkable behaviour under test it seems improbable that this circuit missed any significant fraction of true coincidences. Any failure to detect true coincidences must be the result of the failure of the GM tubes to have a quantum efficiency of 100 per cent. Two distinct methods were used to test for coincidences in the arrival of quanta.

(A) *The first test* for coincidence was made by varying the X-ray intensity (with the crystal in the zero position) and examining the constancy of  $t$  and  $\bar{E}$  as given by equations (6) and (9). The X-ray intensity was changed by altering the X-ray tube current. The only other changes made between runs were in the positions (angles) of the two receiving arms of the spectrometer. In an effort to receive maximum intensity of the diffracted beam, these were readjusted before each run.

Each run (for a fixed tube current) lasted long enough for the coincidence circuit to record more than a hundred counts. This was always several hours, sometimes more than 24 hours.

*The results of this test* are given in Table I. and figs. 2 and 3. In Table I. are listed the experimental data from fifteen runs, the value of  $t$  as calculated by equation (6) and the value of  $\bar{E}$  as calculated from equation (9).

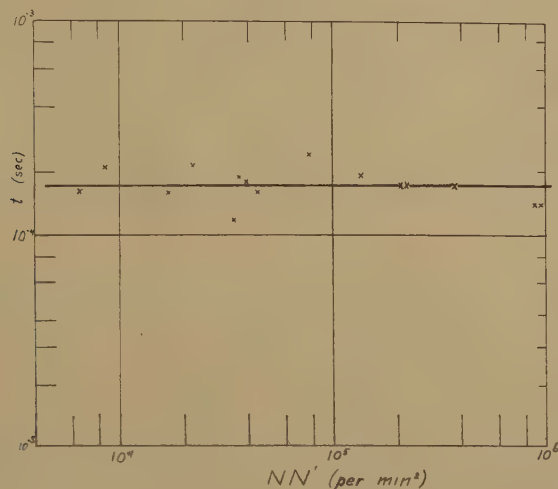
TABLE I.

N (per min.).	N' (per min.).	C (per hr.).	NN' (per min. <sup>2</sup> ).	$t$ (sec.).	$\bar{E}$ (per cent.).
76	86	2.0	$0.65 \times 10^4$	$1.6 \times 10^4$	0.042
100	85	3.5	0.85	2.1	0.064
139	120	5.2	1.67	1.6	0.067
147	150	9.1	2.2	2.1	0.102
171	198	8.2	3.4	1.2	0.073
180	200	13.8	3.6	1.9	0.13
210	186	14.4	3.9	1.8	0.12
190	230	14.5	4.4	1.6	0.12
230	330	35.4	7.6	2.4	0.22
412	320	48.5	13.2	1.9	0.22
430	477	68	20.5	1.7	0.25
418	521	73	22	1.7	0.27
570	640	123	37	1.7	0.34
967	912	249	88	1.4	0.45
1060	888	265	94	1.4	0.45
				Average $t$ $= 1.7 \times 10^{-4}$	

The effect upon resolving time ( $t$ ) and efficiency ( $\bar{E}$ ) of changing the X-ray intensity (N and N'). C is the coincidence counting rate. (Notice the units.)

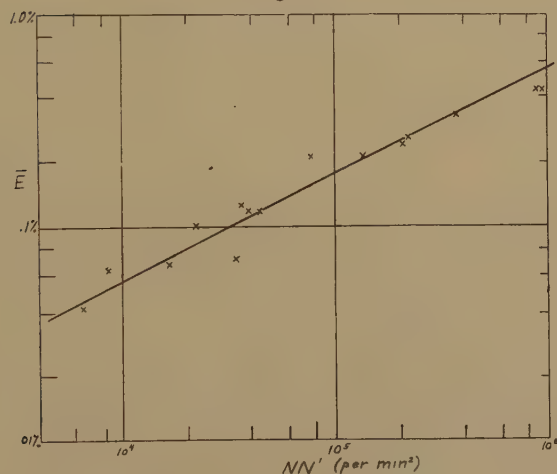
It is at once obvious that the "efficiency,"  $\bar{E}$ , is not constant, so the assumption of complete coherence must be false. Furthermore, the counter "efficiencies," calculated by equation (9), which assumed coherence, not

Fig. 2.



Curve showing the constancy of resolving time ( $t$ ), with change in the counting rate at the two centres,  $N$  and  $N'$ .

Fig. 3.



Curve showing the variation of the apparent efficiency  $\bar{E}$  of the counters, with change in the counting rate at the two counters  $N$  and  $N'$ .

only are not constant but are always *much* less than 1 per cent. These are the values of the efficiency which are required to explain why there

were so few coincidences. The counters are known\* to have an actual quantum efficiency of more than 15 per cent. This discrepancy indicates that the assumption of coherence is incorrect. The value of the resolving time  $t$  fluctuates, but its behaviour appears random with no systematic change as the counting rate increased. This evidence seems to indicate that, aside from experimental and statistical error,  $t$  is constant as was predicted from the quantum picture of diffraction. The average resolving time calculated from the experimental data, by equation (6), is  $1.7 \times 10^{-4}$  sec., which is at least a reasonable figure. Fig. 2 presents this behaviour of  $t$  in graphical form. Fig. 3 shows the behaviour of  $\bar{E}$ .

The above data also serve to illustrate the accuracy which is to be expected in later results where agreement within a factor of two is considered satisfactory.

(B) *The second test* for coincidences could be applied much faster and was much more extensively used. This test was made using an independently obtained value of the resolving time  $t$  and comparing it with the value obtained from a diffraction experiment.

The calibration process consisted of blocking off the X-rays from the counters by means of lead sheets† and then placing near each GM tube a small chunk of uranium ore. Each tube was shielded from the uranium ore placed near the other tube. Each tube then counted bursts from its "own" chunk of uranium ore. The two counters were thus counting bursts from independent sources and only chance coincidences could exist. The values of  $t$  calculated from data obtained in this way are called standard resolving times, and symbolized by  $t_s$ .

Whenever the circuits gave trouble and had to be worked over, the value of  $t_s$  changed, sometimes by as much as a factor of five. It became routine practice to calibrate the apparatus at least twice a week. In this calibration process it was necessary to correct for the fact that the bursts from the uranium were distributed randomly throughout all time, whereas the X-rays were produced from a power source using a half wave rectifier with no filter. Thus the X-rays were "on" less than half the time and the actual counting rate for each counter and for the coincidence circuit was therefore greater than that obtained by dividing the number of counts by the time elapsed. This correction was applied to all data. It is explained in detail in the appendix.

The criterion for randomness in this method is that  $t = t_s$ , where  $t$  is obtained using X-rays and  $t_s$  using uranium. The criterion for coherence between the two diffracted X-ray beams is a reasonable value (cf. Sullivan's value of 15 to 20 per cent.) for the efficiency  $E$  of the GM tubes. This method of distinguishing between randomness and coherence was employed in the following study.

---

\* H. M. Sullivan, *Rev. Sci. Inst.*, xi. p. 356 (1940).

† The X-ray tube was left in operation to insure the same background of electrical disturbance.



To safeguard against the possibility that the zero-position might not have been accurately found, the crystal was rotated in small ( $0.05^\circ$ ) steps about horizontal and vertical axes perpendicular to the incident X-ray beam. After each such rotation a test was made for coherence and randomness. Thus a great variety of angles near the zero position was examined. The results are shown in Tables II. and III. It will be noticed that the crystal was rotated through a greater angle about the vertical axis than about the horizontal. Since the spectrometer was more flexible about the vertical axis it was assumed it was also more subject to error in this direction. It seems improbable that the zero position should have been in error by more than the spread of these tables. As is seen in Table II. the ratio  $t/t_s$  is of the order of magnitude of unity. It is usually between 0.5 and 2.0. It is not sensitive to small changes in

TABLE I.

$\phi \backslash \theta$	-0.25	-0.20	-0.15	-0.10	-0.05	0.0	0.05	0.10	0.15	0.20	0.25
0.10		0.73	2.3 <sup>b</sup>	0.90	1.6	0.12	1.0	1.0	0.40	0.43	
0.05	0.95	0.23 <sup>a</sup>	0.70	0.47	1.3	0.79	1.3	1.1	0.67	1.2	0.98
0.0	1.3	1.2	0.76	0.67	1.0	0.82	1.2	3.0 <sup>c</sup>	1.4	1.5	0.79
-0.05	0.69	0.74	0.83	1.3	1.2	0.78	0.84	1.1	1.0	0.97	0.95
-0.10		1.1	0.90	0.67	0.75	0.90	0.93	1.2	1.1	0.87	

The experimental values of the test ratio  $t/t_s$  for various orientations of the crystal.  $\theta$  is the angular displacement about a vertical axis,  $\phi$  the displacement about a horizontal axis perpendicular to the X-ray beam.

*a, b, c.* A repeat run on each of these values gave 1.2, 0.80 and 1.1 respectively. The original data are left in the table for the sake of the record.

TABLE III.

$\phi \backslash \theta$	-0.25	-0.20	-0.15	-0.10	-0.05	0.0	0.05	0.10	0.15	0.20	0.25
0.10		0.17	0.06	0.06	0.20	0.38	0.025	0.23	0.16	0.10	
0.05	0.02	0.12	0.09	0.20	0.06	0.16	0.24	0.13	0.07	0.06	0.23
0.0	0.12	0.08	0.05	0.06	0.01	0.01	0.02	0.01	0.01	0.12	0.24
-0.05	0.07	0.04	0.06	0.10	0.07	0.05	0.04	0.12	0.06	0.08	0.16
-0.10		0.09	0.06	0.04	0.05	0.20	0.20	0.38	0.06	0.01	

The observed "efficiencies" of the counters (per cent.) for various orientations of the crystal.  $\theta$  is the angular displacement about a vertical axis,  $\phi$  the displacement about a horizontal axis perpendicular to the X-ray beam.

the setting of the crystal. The low efficiency ( $\bar{E}$  in Table III.), which would have to be assumed for the GM tubes in order to account for the observed numbers of coincident counts on the basis of coherent diffraction, is not even of the same order of magnitude as that found experimentally by Sullivan\*. Sullivan found 15 per cent. as the quantum efficiency of the counters used in this laboratory in 1940. The counters now in use are "much more sensitive"†, so that the discrepancy between the efficiency of the GM tubes and the values given in Table III. is at least a factor of 100. This examination of Tables II. and III. has revealed no evidence of any coincidences that cannot be attributed to chance.

### 5. Conclusions.

An extensive search has been made for coherence between the two X-ray beams symmetrically diffracted from the (130) and ( $\bar{1}\bar{3}0$ ) planes of NaCl. No evidence for such coherence has been found. If any portion of the diffracted X-rays is coherent, that portion is too small to be detected by the method and apparatus used in this experiment. If as much as 5 per cent. of the X-rays had been diffracted coherently, that fact could have been detected easily—probably 2 per cent. could have been detected. One is tempted to conclude that the diffraction is entirely incoherent and random as predicted by the quantum picture.

### 6. Acknowledgments.

The author wishes to express his thanks to Dr W. P. Davey for suggesting the problem and for his helpful criticisms. He wishes to thank also the Society of Sigma Xi for a Grant-in-Aid of Research, which made available part of the apparatus.

### Appendix.

It is the purpose of this appendix to present a method whereby the experimental data may be corrected for the fact that the X-ray tube is powered from a half wave rectifier with no filter. It is not practical to filter a rectifier supplying 10 to 20 ma. at 35 to 40 kv.

In the experiment which has been described the GM tubes are placed in position to receive X-rays "reflected" by the (130) and ( $\bar{1}\bar{3}0$ ) planes. The X-rays are incident along a line normal to the (100) planes. Under these conditions the Bragg angle is  $\arctan 1/3$ . From this information and the grating space of NaCl (2.814 A.U.), Bragg's law gives the wave-length of the diffracted X-rays as 0.563 A.U.

This is not a line in the spectrum of the target (Mo), but is part of the continuum. The minimum tube potential for which this wave-length may be produced is 22 kv. The tube was operating from a transformer

---

\* H. M. Sullivan, *loc. cit.*

† Davey, *et al.*, *loc. cit.*



equivalent except in the special case for which  $f(V)$  is a constant, *i. e.* independent of  $V$  (or  $\omega$ ). This is the case in which the X-rays are replaced by bursts from the uranium ore, the behaviour of which were independent of the X-ray tube voltage.

For any other case the two are not equivalent. The coincident circuit records  $\bar{C}$  as given by equation (A 4). The experimental values of the resolving time  $t$  are first obtained from equation (6) which for a fluctuating X-ray intensity is represented by equation (A 6). The value of  $t$  must then be corrected by a factor  $C'/\bar{C}$ .

The value of this ratio  $C'/\bar{C}$  will now be evaluated for two cases which might approximate the correct form for  $f(V)$ . First consider the case in which

$$\left. \begin{aligned} N=f(V)=0, & \quad V < 22 \text{ kv.}, \\ N=f(V)=K(V-22), & \quad V \geq 22 \text{ kv.} \end{aligned} \right\} \quad \text{. . . . . (A 7)}$$

If the X-ray intensity is represented by this function of the potential, then equation (A.4) and (A.6) give

$$C'/\bar{C}=0.248. \quad \text{. . . . . (A 8)}$$

If the intensity is represented by

$$\left. \begin{aligned} N=f(V)=0, & \quad V < 22 \text{ KV}, \\ N=f(V)=K'(V-22)^2, & \quad V \geq 22 \text{ KV}. \end{aligned} \right\} \quad \text{. . . . . (A 9)}$$

then the value of the ratio

$$C'/\bar{C}=0.201 \quad \text{. . . . . (A 10)}$$

Kulenkampff\* has observed experimentally that the function defined by equation (A 7) is approximately correct. Kramers† has given a theoretical derivation, the results of which indicate that equations (A 7) represent a good approximation of the facts.

The values of resolving time  $t$  which are presented in this paper, have been calculated by equation (6) and then multiplied by  $C'/\bar{C}$  ( $=0.248$ ), as given in equation (A 8).

#### *Note.*

For lines in the characteristic spectrum of the target, equations of the form (A 9) would represent a better approximation of the intensity voltage relation ‡.

\* H. Kulenkampff, *Ann. d. Physik*, lxi. p. 548 (1922).

† H. A. Kramers, *Phil. Mag.* xli. p. 836 (1923); also D. L. Webster, *Phys. Rev.* ix. p. 220 (1917).

‡ Sullivan, *loc. cit.*; Webster-Clark, *Proc. Nat. Acad. Sci. U.S.A.* iii. p. 181 (1917).



XXXV. *The Photoelectric Mechanism of Selenium Barrier Layer Elements.*

AV ARNE ELD SANDSTRÖM \*.

[Received July 25, 1945.]

## Introduction.

IN selenium photo-elements, exposure to light occasions a remarkable change, consisting of a lowering of the barrier-layer resistance of the unilluminated element†. After an exposure the resistance increases with time. Measurements on the capacity between the transparent electrode and the selenium layer revealed the fact that this capacity increases when the resistance decreases, and vice versa‡. The model, displayed in fig. 1, was used for the calculations. It was found to obtain for frequencies between 300 and 800 cycles per second.  $C_b$  is the capacity,  $r_b$  the resistance over the barrier layer, and  $C_e$  the capacity between the electrodes. For convenience the electromotive force is not indicated in the diagram. There ought to be a capacity parallel to the resistance  $r_e$  of the selenium layer, too, but this capacity is of no importance where low frequencies are concerned§. The relationship between the capacity and the resistance of the barrier layer turned out to be expressed by the equation

$$C_b = k \cdot \frac{1}{r_b} + b. \quad (1)$$

According to the fact that the changes are not evenly distributed over the whole barrier layer, this equation can be written in a form which shows the significance of the constant  $b$ :

$$C_b = k \cdot \frac{1}{r_b} + C_0 - k \cdot \frac{1}{r_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

$r_0$  is the resistance of that part of the barrier layer, which remains unchanged.  $C_0$  is the corresponding capacity. As could be found from the high values of the constant  $b$ , not only those parts of the element remain unchanged which happen to be screened by the casing, but also the major part of the illuminated area. It was concluded that, by the action of light, metallic crystals were formed at the surface of the barrier layer, thus diminishing the thickness of this layer in certain places. For further particulars the reader is referred to the paper already cited ||.

\* Communicated by Professor N. F. Mott, F.R.S.

† B. Lange, *Phys. Z.* xxxii. p. 850 (1931); 'Die Photoelemente,' p. 118, Leipzig (1936). A. E. Sandström, *Phil. Mag.* (7) xxviii. p. 642 (1939); xxx. p. 428 (1940).

† A. E. Sandström, *Arkiv. f. Matemat, Astr. och Fys*, xxxi. A, No. 2 (1944).

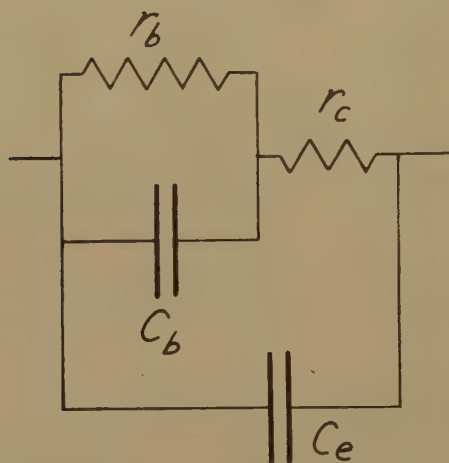
§ *Ibid.* pp. 9 and 29; F. Goos, *ZS. f. Phys.* cxiii. p. 334 (1939).

28 (1944).

Several theories have been proposed for the explanation of the electromotive force of barrier layer elements\*, but most of them were unsatisfactory. In 1939, however, Mott† published a theory accounting both for the electromotive force and for the magnitude and the direction of the photo-current. His theory was applied to the copper-cuprous oxide element.

Between the copper and the oxygen-rich cuprous oxide is a layer of pure oxide, which is an excellent insulator and constitutes a blocking layer‡. The oxygen-rich cuprous oxide is a defect conductor with

Fig. 1.



Model of a photo-voltaic element when unilluminated.  $C_b$  and  $r_b$  are the blocking layer capacity and resistance.  $C_e$  is the capacity between the electrodes.  $r_c$  is usually supposed to be the resistance of the semi-conductor.

empty impurity levels. The levels nearest to the copper capture electrons from the metal, an electric double layer being formed on both sides of the blocking layer. The surface of the metal will be positively charged. The negative charge of the semi-conductor will not be concentrated on the rather ill-defined surface between semi-conductor and blocking layer, but will decrease rapidly with the distance from the metal. Most of the electrons raised from the full bands to the empty conduction levels in the blocking layer will be driven to the metal by the electrostatic field of the double layer. To neutralize these electrons, an equal number of positive holes has to flow across from the oxide to the metal. This is prevented, however, by the direction of the field in the blocking layer and

\* B. Lange, 'Die Photoelemente,' pp. 26-59, Leipzig (1936).

† N. F. Mott, Proc. Roy. Soc. (A), clxxi. p. 281 (1939); N. F. Mott and R. W. Gurney, 'Electronic Processes in Ionic Crystals,' pp. 192-196, Oxford (1940).

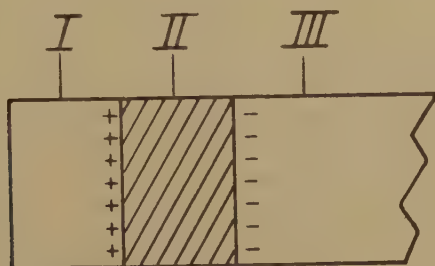
‡ Hitherto, the author has used the expression "barrier layer," which hereafter will be changed to "blocking layer" in accordance with Mott.

thus a potential difference results between the oxide and the metal. The oxide will constitute the positive pole of the element, and the metal the negative pole. When an outer circuit is added, the electrons will flow through this circuit from the metal to the oxide and, naturally, the current in the opposite direction.

*An Application of Mott's Theory to Selenium Elements.*

In the case of selenium photoelements, thanks to its peculiar reaction to illumination, metallic selenium takes the place of the semi-conductor. In ordinary semi-conductors the photo-conductivity is due to impurities or defects in the crystal lattice \*. As regards selenium, however, photo-

Fig. 2.



To the explanation of the electromotive force in photo-voltaic elements. I is the transparent electrode, II is the blocking layer, and III the semi-conductor.

conductivity is a characteristic of the pure element. In this respect it stands alone amongst elements as well as compounds.

Normally there are no electrons in the conduction bands of metallic selenium, the electrons being raised to those levels by the action of light. Therefore, we may assume selenium to be a defect conductor. This is confirmed by the fact that, in the case of selenium photo-voltaic elements, the photo-current flows from the selenium to the transparent gold or platinum electrode. Thus, the selenium has to be positively charged, as is the semi-conductor in the case of copper-cuprous oxide elements. The conditions are illustrated in fig. 2.

It is an interesting feature of Mott's theory that the electromotive force does not depend on any rectifying properties of the element. The only condition to be fulfilled is that the blocking layer has to be highly insulating. This explains the fact that there are selenium elements, which show a considerable photo-electric effect without being rectifying. We will also bear in mind that all blocking layers are non-rectifying for very low voltages, at least up to 10 or 20 millivolts. The photo-voltages are sometimes of a magnitude below this.

\* A. H. Wilson, Proc. Roy. Soc. (A), cxxxiii. p. 458 (1931); cxxxiv. p. 277 (1932); 'Theory of Metals,' Cambridge (1936); 'Semi-conductors and Metals,' Cambridge (1939).

Mott's theory demands that an electron is brought into one of the conduction bands of the blocking layer. However, it is not necessary for the electron to be freed from an atom in the blocking layer itself. It can also be produced in the semi-conductor, if it only gets sufficient initial energy to pass the potential barrier at the contact between semi-conductor and blocking layer. The wave-length limit, however, has to be correspondingly shorter, when the photo-effect takes place in the semi-conductor. Especially, as regards the region of X-rays, it is probable that the electron source is partly to be found in the selenium layer.

#### *The Changes in the Blocking Layer.*

As regards selenium photoelements, the chemical composition of the blocking layer has never been finally ascertained. The best elements are supposed to be formed, when the transparent electrode is applied by cathode sputtering in an oxygen atmosphere. It is probable, therefore, that the blocking layer consists of selenium dioxide. We must not forget, however, that it might consist of a selenide, a selenite or a selenate with the metal of the transparent electrode as the positive ion. As will be explained later, the following discussion is valid in either case. For convenience we will assume the blocking layer to consist of selenium dioxide.

As already mentioned, from experimental evidence it was possible to conclude that the changes of the resistance and the capacity of the blocking layer, when illuminated, are due to a decrease of the thickness of the blocking layer itself. This change is not spread out evenly over the whole insulator but restricted to certain areas. The decrease of the thickness has to be explained by assuming metal to be precipitated either at the semi-conductor or at the transparent electrode. This cannot be the result of any simple electrolysis, as is shown by the fact that it is not possible to produce the change of resistance in question only by putting a voltage on the electrodes. Illumination is essential for the decrease of resistance to take place. Therefore, we will assume that, through the action of light, there are ionized selenium atoms in the blocking layer. These positively charged atoms will then travel in the direction of the negative side of the electric double layer, that is, to the selenium. Arriving at the boundary, these positive ions will catch electrons and build up metallic crystals, thus diminishing the distance over the blocking layer. According to our experience, these new crystals do not cover the whole illuminated area, which can be explained by the fact that crystals cannot form, where the oxide is in immediate contact with the selenium\*. In those parts, however, where there are interstices, even if they are very small, new crystals are able to grow without hindrance. This also explains, why the changes always are the same, or nearly the same, at every new illumination.

---

\* For details on photo-chemical processes, compare N. F. Mott and R. W. Gurney, *ibid.* p. 227.



It is interesting to note that the process is reversible. The blocking layer does not only return to its former condition, when the element is restored to darkness, but even during the illumination, seemingly, a diffusion of metal ions takes place back into the blocking layer. Experiments have shown \* that the final resistance decreases with the time of exposure, as well as with the intensity of the light. If the latter is low, the resistance will reach a constant value after a short exposure already †. The higher the illumination, the lower this constant resistance will be. Further, the fact that, after an exposure, the blocking layer slowly returns to its former condition indicates the instability of the freshly formed accumulations of metal atoms.

When the positively charged atoms reach the selenium surface, they meet a stream of electrons travelling in the opposite direction to replace those lost by photo-effect. Then it is possible that the newly formed small crystals are torn away in the same manner as in the case of cathode sputtering. They will have a negative charge and are drawn through the crystal lattice of the blocking layer in the direction of the transparent electrode. Probably, very soon the negative charges will be released and then we will have either vagabonding neutral selenium atoms or selenium ions striving to reunite with oxygen. If the stability of the new crystals decreases with increasing size, after a time as many atoms will leave the surface of the selenium as those arriving there and the conditions have become stable. This explains why the resistance becomes constant after a finite time of exposure, as well as the fact that this constant value depends on the intensity of the incident light.

It is also possible that the crystals, which are formed on the surface between the selenium layer and the dioxide, are too small to exist, when charged. Then they will be torn to pieces and thrown back into the blocking layer, even after the end of the exposure. The probability that they are built on to the original selenium crystals rather loosely, makes this explanation the more possible.

If, through the photo-effect, the selenium dioxide is being dissociated, we will also get positive oxygen ions. These are drawn to the negative side of the double layer. The free oxygen cannot leave the photoelement, which is, usually, covered by a thin sheet of cellulose acetate. After the end of an exposure, the oxygen ions will strive to reunite with the selenium ions and, therefore, they diffundate back into the blocking layer.

We shall return later to some of the remarks above.

As regards experimental evidence of the wandering of selenium atoms in the direction of the negative side of the double layer, besides those already cited, the following seem to be of interest. The photo-electromotive force between the selenium and the transparent electrode will

---

\* A. E. Sandström, *Phil. Mag.* (7) xxviii, p. 642 (1939).

† A. E. Sandström, *Arkiv f. Matemat, Astr. och Fys*, xxxi. A, No. 2, p. 26, fig. 13 (1944).

only slightly affect the field in the blocking layer. However, if a several volts higher voltage is applied between the electrodes, it might be possible to depress the field in the blocking layer by giving the selenium a high positive voltage against the other electrode. In the same way, by putting a negative voltage on the selenium, we ought to be able to increase the field in the blocking layer. Really, the change of resistance was found to be three or four times as large in the latter case as in the former. The voltages were kept between 3 and 15 volts. According to this experiment, too, the changes seem to be due to a transport of metal atoms to the selenium layer. If the voltage was increased to about 20 volts, no difference could be observed as to the resulting resistance, when the direction of the field was reversed. In this case, probably, the field is so strong as to make the primary field of the double layer negligible. The charged atoms will be transported to one electrode as easily as to the other.

For comparison purposes we shall make some numerical calculations on the three elements, used in an earlier investigation\*.

The thickness of the blocking layer can be calculated from the blocking layer capacity  $C_b$  of the unilluminated element

$$C_b = \epsilon \frac{S}{4\pi d} \quad (2)$$

$S$  is the total area of the photoelement,  $\epsilon$  the dielectric constant of the blocking layer and  $d$  the thickness to be calculated. The dielectric constant will be of the order of magnitude of 10. The values of the blocking layer thickness are given in the fifth column of Table I.

As regards the blocking layer resistance, putting  $\rho$  for the resistivity, we have the relation

$$r_b = \rho \frac{d}{S} \quad (3)$$

Having obtained the thickness  $d$  from equation (2), we are now able to calculate the resistivity, which we will define as the resistance between two opposite surfaces of a cubic body with its edge equalling 1 cm. The values for the three elements in question are also to be found in Table I.

Evidently, the blocking layer consists of a good insulator. However, the differences between the three elements are considerable, both as regards the thickness and the resistivity of the blocking layer, showing that much is left to fate in the manufacturing of selenium elements.

Equations (2) and (3) give us the relation  $r_b C_b = \frac{\epsilon \rho}{4\pi}$ . If  $r_0$  and  $C_0$  denote the resistance and the capacity of the unchangeable part of the blocking layer, in combination with equations (1) and (1a), we get

$$\begin{cases} r_0 C_0 = \frac{\epsilon \rho}{4\pi}, \\ C_0 - \frac{k}{r_0} = b. \end{cases}$$

---

\* A. E. Sandström, *ibid.*

Eliminating  $r_0$  we find

$$C_0 = \frac{b}{1 - \frac{4\pi k}{\epsilon\rho}}$$

Using the values of  $\rho$  from Table I., and remembering  $k$  to be \* of the order of magnitude of 500,  $\frac{4\pi k}{\epsilon\rho}$  will be of the order of magnitude of  $10^{-7}$ .

Thus  $C_0 = b$ . . . . . (4)

TABLE I.

Element.	Blocking layer capacity $C_b'$ of the unexposed element. $\mu\text{F.}$	Blocking layer resistance $r_b'$ of the unexposed element. Ohms.	Total area of element. $\text{cm.}^2$	$a$ $10^{-4} \text{ cm.}$	$\rho$ $10^9 \text{ ohms.}$
E I	0.120	40000	4.9	0.31	6.4
E II	0.105	34000	5.0	0.40	4.3
E III	0.270	9000	19.2	0.63	2.8

As  $b$  can be determined \*, it is possible to calculate the ratio  $C_0/C_b'$ ,  $C_b'$  being the blocking layer capacity of the unexposed element. From the values of this ratio (Table II.) can be seen that only a very small part of

TABLE II.

Element.	$Cb=r$ . $\mu\text{F.}$	$C_0/C_b'$ .	$r_0$ . Ohms.	Lowest blocking layer resistance after exposure. Ohms.	$r_1$ . Ohms.	$d_x$ . $10^{-6} \text{ cm.}$
E I	0.117	0.97	41000	5000	5700	0.13
E II	0.089	0.81	43000	12000	16600	3.9
E III	0.200	0.74	12400	2400	3000	5.3

the blocking layer partakes in the changes. As regards Element I 97 per cent. of the blocking layer remains unchanged. This is in complete agreement with the fact that the change of resistance is very large, and the change of capacity very small. As the thickness and resistivity of the blocking layer are known, we are able to calculate the resistance  $r_0$  of the unchangeable part of the blocking layer. From the resistance after an exposure, we are then able to calculate the lowest resistance of the changeable part and, accordingly, also the new thickness  $d_x$  of that part of the blocking layer, which changes under the influence of illumination.

Taking the lowest observed values of  $r_1$ , we obtain the  $d_x$ -values in the seventh column of Table II. Naturally, the blocking layer does not

\* A. E. Sandström, *ibid.* pp. 20-23.



decrease in thickness to the same amount in every place. Therefore, the thickness  $d_x$  has to be regarded as an average. In many places the newly formed metal crystals will penetrate the blocking layer, even to a farther extent than shown by the values of  $d_x$  and  $d-d_x$  (Table III.).

Selenium photoelements are known to change their resistance during storage. An increase of resistance means that they are not as yet fully restored after their latest exposure. Later on their resistance often decreases considerably. Usually this is explained by assuming conducting bridges to form across the blocking layer. The correctness of this assumption is proved by the fact that these "bridges" can be burnt off by sending an electric shock through the element. The above calculations leading to the extremely small values of  $d_x$ , especially as regards E. I., is a further proof of the reality of these bridges. The spontaneous forming of such conducting bridges during a prolonged storage can be the result of accidental ionization by heat. It seems possible, too, that slowly growing accumulations of metal are more stable than those formed during illumination. When the bridges are destroyed by electric shocks, *e. g.* heating, the atoms will diffuse back into the crystal lattice of the blocking layer.

Now we also get a clearer view as regards the instability of the freshly formed crystals during illumination. As these are appearing as thin threads in cracks in or between the crystals of the blocking layer, they will be very susceptible to any actions of electric charges.

#### *On the Composition of the Blocking Layer and the Source of the Secondary Photo-Electric Phenomena.*

Hitherto we have assumed the blocking layer to consist of selenium dioxide, although it is possible, also, that it consists of a compound with the selenium as a part of the negative ion. The metal of the transparent electrode will then constitute the positive ion. It is also possible that the blocking layer will be intermixed with atoms from this electrode even if the main part of it consists of selenium dioxide. Anyhow, if the transported metal atoms are of the same kind as those of the transparent electrode, by illumination we will, in some places, get the same metal on both sides of the blocking layer. Now, if the material is the same on both sides of an insulator, naturally, no electric double layer will be formed and no electromotive force will come into existence. According to our experimental evidence, however, only a small part of the surface will be covered and, the field in the insulator will not be seriously influenced. For confirmation an element was exposed for a considerable time with a rather high positive voltage on the selenium, but without any diminishing effect whatever on the photo-electric effect. As follows from the above arguments, however, no conclusions can be drawn from this result as to the nature of the compound forming the blocking layer. We will only remember that our explanation of the changes in the blocking layer are not affected by the question of the blocking layer material.

Knowing the precipitated volume of metal, we are able to calculate the number of atoms participating in the process of forming new crystals



on the surface of the selenium layer. Once more, assuming the precipitated metal to be selenium, we get the values in Table III. If the metal is platinum or gold we will get twice as many atoms, but the order of magnitude will be the same. The sixth column of Table III. gives the number of precipitated atoms per cubic centimetre of the blocking layer material.

TABLE III.

Element.	Changeable blocking layer area. cm. <sup>2</sup>	$d-d_x$ . $10^{-4}$ cm.	Volume of precipitated metal. cm. <sup>3</sup>	Number of atoms.	Number of precipitated atoms per cm. <sup>3</sup> of blocking layer material.
E I	0.15	0.31	$4.6 \times 10^{-6}$	$1.5 \times 10^{17}$	$1.0 \times 10^{21}$
E II	0.45	0.36	$3.6 \times 10^{-5}$	$1.2 \times 10^{18}$	$6.0 \times 10^{21}$
E III	5.0	0.58	$2.9 \times 10^{-4}$	$9.5 \times 10^{18}$	$7.7 \times 10^{21}$

As already mentioned, the blocking layer consists of a good insulator. Its insulating properties are considerably heightened with decreasing temperature. For one element the resistance at room temperature was found to be 47000 ohms, but at liquid air temperature 120000 ohms. It is possible that the existing conductivity depends on impurities and lattice defects. The impurities can either consist of too rich an abundance of selenium atoms in the blocking layer, or an equally rich abundance of metallic atoms from the transparent electrode. Naturally, we may expect this excess of metal atoms to constitute the main part, if not the whole, of the flow of ions building up the new crystals from the border between selenium and the blocking layer. Table III. shows these atoms to be as many as  $10^{21}$  per cubic centimetre of the blocking layer. This is  $10^4$  times as many as those observed in connection with the investigations on the properties of halide crystals\*. Nevertheless, their number is still small as compared with the number of atoms in the crystal lattice of the blocking layer. Thus, it seems probable that the observed changes of resistance and capacity are due to a surplus of metal atoms in the blocking layer. It ought to be possible to better the properties of selenium elements by applying the blocking layer in such a way as to keep it as pure as possible. We have to remember, however, that we do not know to what extent impurities are necessary for a large production of photo-electrons.

*On the Resistance  $r_0$  in Series with the Electromotive Force.*

It has been found that the value of the resistance in series with the electromotive force is many times higher, when measured in the alternating current circuit, than when measured in the direct current circuit†. In the latter case the resistance is taken from the approximately constant part of the rectifier curve of the unilluminated element. The resistance

\* R. W. Pohl, *Phys. Z.* xxxix. p. 36 (1938).

† A. E. Sandström, *Arkiv f. Matemat, Astr. och Fysik*, xxxi. A, No. 2 (1944).

for voltages as high as 20 to 40 volts has been considered as due to the selenium layer only. For one element the author found  $r_c=800$  ohms when measured in this way but  $r_c=3150$  ohms when measured with alternating current. It is possible, however, that the low value in the former case is due to a discharge across the blocking layer, a pressure of 20 volts between the electrodes giving a drop of 640000 volts per cm. in the blocking layer.

A closer study of the nature of the resistance  $r_c$  in connection with our conception, given above, of the photo-electric mechanism reveals some new facts. When the electrodes are connected electrically the electrons do not only flow from the transparent electrode to the selenium layer, but the tendency towards an equalization between the negative and positive charges drives them on until they meet those positive holes, which they are to neutralize. It is possible, and even very probable, that the positive holes in the blocking layer move towards the negative side of the electric double layer, but it does not follow that they move as speedily as the electrons. Therefore, it is possible that the act of neutralization takes place only in part in the selenium layer. Some electrons may have to travel a shorter or longer distance through the blocking layer to meet positive holes. Thus, we get a set of parallel currents, some in leads with a lower, and some in leads with a higher resistance.  $r_c$  will then represent a combination resistance, into which enters a part of the blocking layer resistance. Therefore, the true resistance  $r_c$  during illumination might be quite different from that measured by the aid of an enforced voltage between the electrodes. Discussions will be postponed, however, until further measurements are available.

#### *Summary.*

1. Mott's theory of the electromotive force of oxide photoelements is applied to the selenium element.

2. Illumination causes an increase of the blocking layer capacity as well as a decrease of the blocking layer resistance of selenium photoelements. In an earlier paper the author explained this by assuming metal crystals to form in the border between the selenium and the blocking layer, thus making the blocking layer thinner. In the present paper this idea is being followed up with a view to some consequences following from Mott's theory. An explanation is given as to how such an accumulation of material is possible.

3. Former measurements on three selenium photoelements are used for numerical calculations confirming the discussion. The well known fact of conducting bridges spontaneously forming over the blocking layer during storage in darkness, is explained.

4. Most of the precipitated atoms are impurities in the blocking layer. They consist either of selenium or of the metal of the transparent electrode.

5. Some notes are given on the resistance in series with the electromotive force.

XXXVI. *Note on the Empirical Relations between Chemical Constitution and Thermal Expansion Coefficient of Organic Liquids* \*.

By GUNNAR STEENSHOLT.

[Received April 15, 1946.]

FOR several important physical properties and characteristic constants of organic compounds a large body of experimental material has been systematically accumulated in order to study their dependence on chemical constitution, the influence of various functional groups in the molecule, and so on. In this connection mention may be made of the data on melting and boiling points, di-pole moments, etc. Undoubtedly, however, much remain to be done in this field. Thus, so far no discussion seems to have been given in the literature of the dependence of the thermal expansion coefficient of organic liquids on their molecular structure †, in spite of the fact that a considerable amount of experimental material has been brought together. Only two papers, one by Ellis and Reid ‡ and the other by Walden, Ulich and Birr §, contain a discussion of some particular classes of compounds, namely certain mercaptans resp. ammonium picrates. Beyond this the present author does not know of any attempt to discuss more generally the problem just mentioned, though it is perhaps not unlikely that some of the more pronounced regularities, which will be discussed below, have been known to experimental physicists working in this field.

It should be stressed at once that there is a great need for further and above all more systematic experimental data on the thermal expansion coefficient of liquids. This, it is hoped, will be brought out in the following discussion. It is therefore clear that any attempt at the present time at a co-ordination of the empirical facts must necessarily be of a qualitative and in some respects also of a preliminary nature. Nevertheless, the author believes that an investigation of this kind should be carried out, since it may serve the double purpose of indicating the direction of further experimental work, and at the same time of providing a certain fundamental framework for the theoretical physicist, whose task it will be some day to interpret and explain the empirical rules on the basis of a general quantum mechanical theory.

The experimental material on the thermal expansion coefficient of organic liquids is very scattered in the literature, and is partly to be found

---

\* Communicated by the Author.

† A definite statement to this effect is made by W. Hückel, 'Theoretische Grundlagen der organischen Chemie,' vol. ii. p. 28 (1941).

‡ J. Am. Chem. Soc. vol. liv. pp. 1674-1687 (1932).

§ ZS. physikal. Chem. vol. cxxx. p. 495 (1937).



in rather inaccessible journals, but a summary in tabular form is to be found in Landolt-Bornstein's tables, 'Hauptwerk,' p. 1224, and 'Ergänzungswerk,' I. p. 686, II. p. 1156, and III. p. 2222. The reader is therefore referred to this source, where also references to the original literature are to be found.

On first surveying the data the question may naturally be asked whether we are at all permitted to compare the thermal expansion coefficients of two arbitrarily given organic liquids as found in the literature, the main objection to such a direct comparison being the fact that the coefficients are very often measured at different temperatures or in different temperature intervals. However, experience shows that the variations of the thermal expansion coefficients with temperature are small, at any rate in the intervals considered here, and we may therefore feel justified in comparing the coefficients of two liquids in a straightforward way.

We shall first consider the aromatic compounds. On examining the experimental material one is struck by the following fact concerning the thermal expansion coefficient of benzene and its derivatives: *If a substitution is made in benzene, the thermal expansion coefficient of the new compound is lower than that of benzene itself.* For benzene, two values of the thermal expansion coefficient are given in the literature, namely

$$0.0_{2}1229 (0-30^{\circ}) *$$

and

$$0.0_{2}106 (0-30^{\circ}) \dagger \ddagger.$$

Which one is the more correct we cannot say, but accepting the mean of these two values as the correct one for benzene, we find in the tables only one compound which perhaps constitutes an exception to this rule, namely fluorobenzene, with an expansion coefficient  $\alpha=0.0_{2}116 (0-30^{\circ})$ . An experimental re-investigation of this compound would be important.

For toluene  $\alpha=0.0_{2}106 (0-30^{\circ})$ , and there is some indication that for benzene the introduction of a methyl group produces only a particularly small decrease in the thermal expansion coefficient. However, further discussion of which groups are the most efficient in decreasing the thermal expansion coefficient of benzene must be left out at the present time.

As regards the hydroaromatic compound cyclohexane there are some indications of the validity of a similar rule. The evidence is admittedly rather scanty and is summarized in Table I.

A further striking fact noticed at once in Table I. is that the thermal expansion coefficient is decreasing in the *o*-, *m*- and *p*-sequence, both for cyclohexanol and cyclohexanon. When more experimental material has become available, this rule may possibly be generalized to some extent.

\* G. Tammann and W. Hirschberg, *ZS. physikal. Chem.* xiii. p. 543 (1894).

† J. Timmermans and F. Martin, *Journ. chim. physique*, vol. xxiii. p. 747 (1926); *cfr.* also Br. Lachowicz, *Ber. deutsch. chem. Ges.* vol. xxi. p. 2206 (1888), where values from 0.0<sub>2</sub>114 to 0.0<sub>2</sub>146 are given.

‡ Here and in the sequel the temperature range to which the measurements refer, is indicated in brackets.



TABLE I.

Cyclohexane .....	0.0 <sub>2</sub> 120 (15-30)
Cyclohexanol .....	0.0 <sub>2</sub> 0825 (39.1-65.9)
Cyclohexanon .....	0.0 <sub>2</sub> 09140 (12.2-29.6)
<i>o</i> -Cyclohexanol .....	0.0 <sub>2</sub> 09847 (39.1-65.9)
<i>m</i> -Cyclohexanol .....	0.0 <sub>2</sub> 08700 (—)
<i>p</i> -Cyclohexanol .....	0.0 <sub>2</sub> 08566 (—)
<i>o</i> -Methylcyclohexanon .....	0.0 <sub>2</sub> 09334 (15.1-39.1)
<i>m</i> -Methylcyclohexanon .....	0.0 <sub>2</sub> 09128 (17.3-39.1)
<i>p</i> -Methylcyclohexanon .....	0.0 <sub>2</sub> 08920 (—)
Methylcyclohexane .....	0.0 <sub>2</sub> 118 (0-30)

Also as regards aliphatic compounds general rules are difficult to formulate at the present time, owing to the incompleteness of the data, but some rather striking regularities can, nevertheless, be brought out. For this purpose we summarize in the following tables the values of the thermal expansion coefficients for some classes of compounds, namely some saturated fatty acids, their nitriles and some simple esters.

TABLE II.—Acids.

Formic acid .....	0.0 <sub>2</sub> 102 (15-30)
Acetic acid .....	0.0 <sub>2</sub> 107 (30)
Propionic acid .....	0.0 <sub>2</sub> 109 (0-30)
„ <i>n</i> .....	0.0 <sub>2</sub> 104 (0-30)
Butyric acid <i>i</i> .....	0.0 <sub>2</sub> 108 (0-30)
„ <i>n</i> .....	0.0 <sub>2</sub> 094 (0-30)
Valeric acid <i>i</i> .....	0.0 <sub>2</sub> 100 (0-30)
Octylic acid <i>n</i> .....	0.0 <sub>2</sub> 0884 (0-30)
Pelargonic acid .....	0.0 <sub>2</sub> 0848 (0-30)

TABLE III.—Nitriles.

Acetonitrile .....	0.0 <sub>2</sub> 138 (0-30)
Propionitrile .....	0.0 <sub>2</sub> 127 (0-30)
„ <i>n</i> .....	0.0 <sub>2</sub> 0849 (15-30)
Butyronitrile <i>i</i> .....	0.0 <sub>2</sub> 116 (0-30)
„ <i>n</i> .....	0.0 <sub>2</sub> 108 (0-30)
Valeronitrile <i>i</i> .....	0.0 <sub>2</sub> 167 (0-30)
<i>n</i> -Octylnitrile .....	0.0 <sub>2</sub> 0938 (0-30)
Pelargonitrile .....	0.0 <sub>2</sub> 0898 (0-30)

TABLE IV.—Methyl esters.

Methyl formiate .....	0.0 <sub>2</sub> 147 (0-30)
Methyl valerianate .....	0.0 <sub>2</sub> 111 (0-15)

TABLE V.—Ethyl esters.

Ethyl formiate .....	0.0 <sub>2</sub> 141 (0-30)
Ethyl acetate .....	0.0 <sub>2</sub> 135 (0-30)
Ethyl propionate .....	0.0 <sub>2</sub> 127 (0-30)
Ethyl butyrate .....	0.0 <sub>2</sub> 119 (0-30)
<i>n</i> -Octylic acid ethyl ester .....	0.0 <sub>2</sub> 0972 (0-30)
<i>n</i> -Nonylic acid ethyl ester .....	0.0 <sub>2</sub> 0952 (0-30)

TABLE VI.—*n*-butyl esters.

<i>n</i> -Butyl propionate .....	0·0 <sub>2</sub> 116 (0-15)
<i>n</i> -Butyl butyrate .....	0·0 <sub>2</sub> 111 (0-15)
<i>n</i> -Butyl valerianate .....	0·0 <sub>2</sub> 116 (0-15)

TABLE VII.—*n*-amyl esters.

<i>n</i> -Amyl formiate .....	0·0 <sub>2</sub> 109 (0-15)
<i>n</i> -Amyl acetate .....	0·0 <sub>2</sub> 114 (0-15)
<i>n</i> -Amyl propionate .....	0·0 <sub>2</sub> 116 (0-15)
<i>n</i> -Amyl butyrate .....	0·0 <sub>2</sub> 114 (0-15)

On surveying the data contained in these tables we readily establish the following rules for the compounds considered :—

- (a) The nitriles have a higher value of the thermal expansion coefficient than the corresponding fatty acid.
- (b) The esters have a greater value of the thermal expansion coefficient than the corresponding acid.

No simple qualitative relationships seem to exist between the relative magnitude of the thermal expansion coefficients of esters and the corresponding alcohols. To test the generality of these rules we must await further systematic experimental work.

For the alkyl halogenides the experimental values are summarized in the following table :—

TABLE VIII.—Alkyl halogenides.

	Chloride	Bromide	Iodide
Methyl .....	—	—	0·0 <sub>2</sub> 087
Ethyl .....	—	0·0 <sub>2</sub> 142	0·0 <sub>2</sub> 117
„ <i>n</i> .....	0·0 <sub>2</sub> 139	0·0 <sub>2</sub> 145	0·0 <sub>2</sub> 1095
Propyl <i>i</i> .....	0·0 <sub>2</sub> 147	0·0 <sub>2</sub> 128	0·0 <sub>2</sub> 113
„ <i>n</i> .....	0·0 <sub>2</sub> 145	0·0 <sub>2</sub> 115	0·0 <sub>2</sub> 102
Butyl <i>i</i> .....	0·0 <sub>2</sub> 129	0·0 <sub>2</sub> 117	0·0 <sub>2</sub> 104
„ <i>n</i> .....	0·0 <sub>2</sub> 124	0·0 <sub>2</sub> 115	0·0 <sub>2</sub> 102
Amyl <i>i</i> .....	0·0 <sub>2</sub> 129	0·0 <sub>2</sub> 117	0·0 <sub>2</sub> 104

(All values in this table refer to the temperature range 0-30° C.)

The striking feature of this table is of course the decrease in the thermal expansion coefficient in the halogen sequence chloride, bromide, iodide.

Finally, we would like to draw attention to the fact that, considering the normal and iso compounds given in this and the preceding tables, one finds that with the exception of butyl chloride an iso compound always has a higher value of the thermal expansion coefficient than the corresponding normal compound.

All in all, then, as shown above, several empirical rules can be readily established, but much more experimental work is necessary before further progress is possible in this direction.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXXVII. *Interference Fringes due to a Wedge of Air or Liquid between Two Glass Plates, in Theory and Practice* †.

By G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge ‡.

[Received June 4, 1945.]

§ 1. *Introduction.*

LET the opposed surfaces U, V, of two "plane parallel" plates of glass, inclined at 0.001 radian, and separated by air, cut a horizontal plane perpendicular to both—a principal plane—in Ou, Ov. Let one plate bear a mark J on its inner surface in the plane Ouv and 5 cm. from O. The distance between the plates at J is then 0.005 cm., or 85 wave-lengths of sodium light ( $\lambda = 5.893 \times 10^{-5}$  cm.). If an observer look, in the direction of J, through the plates towards a sodium flame, or if he view it by reflexion, he sees vertical interference fringes with a spacing of 0.02946 cm. If U, V be semi-silvered, the fringes are rendered more vigorous. The fringes are not located in the "wedge" unless the incidence of the light be normal. In the case supposed, the fringe corresponding to J is 3.535 cm. from J, when the angle of incidence is  $45^\circ$ . No more than a pin and the method of parallax is needed to locate a fringe (see § 15). When the angle of incidence is changed, the fringes move relative to the plates. When a correction for the finite thickness of the plate between the air "wedge" and the observer is applied, the fringe F, corresponding to J, meets the plane Uuv in a point whose locus is a circle § on OJ as diameter; a more precise statement is made in § 10. The diameter of the circle is thus independent of  $\lambda$ ; the spacing of the fringes is proportioned to  $\lambda$ .

The few text books dealing with the theory consider only a single pair of rays, and they in a principal plane. Both come from a luminous point Q and leave Q in the same direction. One is reflected at both U and V; the other suffers no reflexion. They meet in F at an angle, in the case supposed, of 0.002 radians or 7 minutes. No effective determination of the point of intersection could be expected under these conditions. Actually, however, the two rays are of cardinal importance. They will be called the primal pair of rays.

† Some years ago, Mr. C. H. Westcott arranged an experiment on the location of fringes. His work, of a preliminary nature, instigated the investigations of the present paper.

‡ Communicated by the Author.

§ When the medium forming the wedge has  $\mu > 1$ , the locus is not a circle.

### § 2. *The Optical Problem.*

We have to explain how an air wedge between two plates can form bright and relatively dark fringes, when the source of light is a sodium flame a few cm. from the plates whose available area is 2 cm.<sup>2</sup> or more. When the fringes are "real" they can be received on ground glass. The pattern is comparable with that obtained when the image of a wire grating of say, 1 mm. pitch, backed by a sodium flame, is formed on ground glass by a lens of small aperture. Yet there is no grating to serve as object, no lens to form an image and the aperture is so large that any pin-hole effect is excluded. Further, the quality of the pattern does not depend upon the angle of incidence of the light.

When a lens of small aperture is used, all the rays, which come from an object point on its axis and pass through the lens, meet again in an image point; the optical lengths of the paths of *all* these rays are ideally equal.

With the wedge, in the case of transmission, one and only one ray from any luminous point S passes actually or in direction through any given "real" or "virtual" goal point G, after one reflexion at U and one at V. Another unique ray from S passes actually or in direction through G without suffering reflexion. These *two* are the *only* rays passing from S to G, unless we include rays suffering more than two reflexions; multiple reflexions are considered later.

In another case, one ray reaching G is reflected at U and the other at V. This case may be derived by a simple transformation from that of transmission.

The angle between the two rays is very small—of the order of 0.001 radian. The ratio of the amplitude at G of the disturbance due to the twice reflected ray to that due to the other ray, is less than unity. If the plates be suitably semi-silvered, the ratio is of the order of one-quarter or more. With non-silvered plates, it may be 1/25. When each ray suffers one reflexion, the ratio, for non-silvered plates, is more nearly unity, and such plates, with oblique incidence—say 45°—give a pattern of strong contrasts. The two rays assist or oppose each other according as they agree or differ in phase at G.

It is clear that, for a given goal G, we must, even if we neglect rays suffering more than two reflexions, take account, not merely of the primal pair of rays selected by the text books, but of *every* pair, consisting of a twice reflected ray and its non-reflected partner, which reach G through the available aperture of the plates.

When an air wedge is flanked by two glass plates, or when a wedge with  $\mu > 1$  is or is not flanked by two glass plates, an exact account of the path difference between the members of *every* pair of rays becomes impracticable through complications of analysis, and we are driven, even for the primal rays, to neglect  $\eta^2$  and higher powers of the small angle  $\eta$  of the wedge, and are further restricted to rays making angles of not more than a few degrees with the primal pair. These approximations, however, yield results which are satisfactory in practical work. But now we must



take account of changes of phase. If the phase be changed on reflexion, it is changed by  $\pi$ —a reversal; it is not changed on transmission.

In the ideal case considered in § 3, exact results are easily obtained. Although this case cannot be realized, the theory serves as an introduction and provides a standard with which the approximate results for actual cases can be compared. For his help in the theory of the ideal case, I owe much to my friend, the late Dr. G. T. Bennett, F.R.S. He had for many years put his great mathematical skill at my service.

### § 3. Path Differences in Ideal Case.

In the ideal case, two half-planes U, V, meet in the edge Oz; a plane with Oz as normal cuts them in Ou, Ov (figs. 1, 2). The half-planes do not extend beyond Oz. The medium between U and V is air. The plane Ouv and its parallels are principal planes. The (small) angle uOv is  $\eta$ . The planes can transmit light without bending the rays and can reflect

Fig. 1.

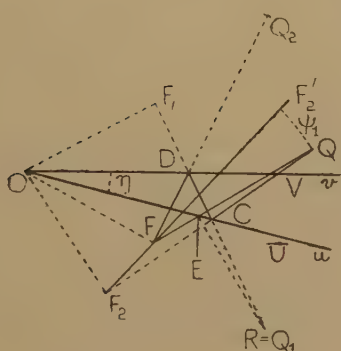
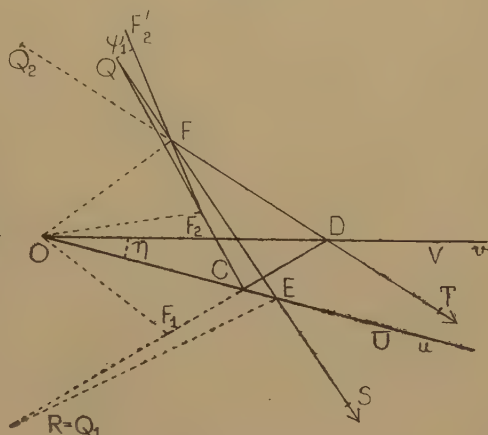


Fig. 2.



light. We suppose that no change of phase occurs at transmission or reflexion. In certain practical cases the wedge has flanking plates, and changes of phase occur; these will be noted in due course.

In the plane Ouv we take a point F at distance  $f$  from O. In fig. 1, Ou lies between F and Ov; in fig. 2, Ov lies between F and Ou. Let  $F_1$  be the image of F in V, and  $F_2$  that of  $F_1$  in U. Because V bisects  $FF_1$  at right angles,  $OF_1 = OF$ ; similarly  $OF_2 = OF_1$ . Hence F,  $F_1$ ,  $F_2$  lie on a circle with centre O and radius  $f$ . The angle  $FF_1F_2$  between the normals  $FF_1$  and  $F_2F_1$  is  $\eta$ , and thus

$$FOF_2 = 2\eta \quad \dots \dots \dots (3.1)$$

Hence

$$F_2F = 2f \sin \eta \quad \dots \dots \dots (3.2)$$

We call  $F_1$  the first, and  $F_2$  the second, image of F.

The straight line  $F_2F$  is of special importance. We call it the first Cardinal Axis of F.

Let Q be a luminous point at distance  $r$  from F, and let FQ make angle  $\psi_1$  with  $F_2F$  produced. In figs. 1, 2 the half-plane V is between Q and U. The point Q is not restricted to be in the principal plane of F. From Q one ray—the direct ray—reaches F along QF; its path length is QF or  $r$ . A second ray—a zigzag of order 1—reaches F after two reflexions, first at C on U, and then at D on V. The points C, D and the length of the zigzag are easily found. Draw  $QF_2$  meeting U in C. Since  $F_1$  is the image of  $F_2$  in U, the ray reflected at C has direction  $CF_1$ ; it meets V in D. Since F is the image of  $F_1$  in V, the ray reflected at D passes through F, actually in fig. 1, or in direction in fig. 2.

By the laws of reflexion,  $DF=DF_1$ ,  $CF_1=CF_2$ , and  $CDF_1$ ,  $QCF_2$  are straight. Hence, if  $[QF]_1$  be the length of the zigzag, we have, in fig. 1,

$$[QF]_1=QC+CD+DF=QC+CD+DF_1=QC+CF_1=QC+CF_2.$$

and thus (3.3)
$$[QF]_1=QF_2.$$

The points Q, C, D in figs. 1, 2 are the projections on the principal plane of the actual points. The projections of QC, CD, DF obey the law of reflexion, and QC, CD, DF are equally inclined to the plane.

Let the length of the general zigzag of order 1 exceed QF by  $\Delta_1^*$ . Then, in fig. 1,

$$\Delta_1^*=[QF]_1-QF=QF_2-QF. \tag{3.4}$$

In fig. 2, we take S on FE and T on FD, with  $FT=FS$ . Then  $\Delta_1^*=[QT]_1-QS$ . For the zigzag,  $[QT]_1=QF_2+FT$ . Hence

$$\Delta_1^*=QF_2+FT-(QF+FS)=QF_2-QF. \tag{3.5}$$

Let R, or  $Q_1$ , be the image of Q in U, and  $Q_2$  that of R in V. Then both RCD and  $Q_2DF$  are straight. Hence

$$\begin{aligned} [QF]_1 &= QC+CD+DF=RC+CD+DF=RD+DF \\ &= Q_2D+DF=Q_2F. \end{aligned} \tag{3.6}$$

Thus (3.7)
$$\Delta_1^*=[QF]_1-QF=Q_2F-QF.$$

This form of  $\Delta_1^*$  is not used in the present paper.

The images  $Q_1, Q_2 \dots$  of Q present themselves more naturally than those of F. But, when  $Q_2F$  is used instead of its equal  $QF_2$ , the problem is unattractive. It is, perhaps, for this reason that the wedge has received little serious attention.

When Q lies on  $F_2F$ , but so that F is between  $F_2$  and Q, we have  $\psi_1=0$ , and  $QF_2-QF=F_2F$ . In this cardinal case, the difference is denoted by  $\Delta_1$ , and by (3.2),

$$\Delta_1=F_2F=2f \sin \eta. \tag{3.8}$$

We have

$$FQ=r, \quad F_2Q=r+\Delta_1^*, \quad F_2F=\Delta_1, \quad F_2FQ=\pi-\psi_1,$$

and (3.9)
$$r+\Delta_1^*=\{r^2+2r\Delta_1 \cos \psi_1+\Delta_1^2\}^{\frac{1}{2}}.$$

Hence  $\Delta_1^*$  depends only on  $\Delta_1, r$  and  $\psi_1$ , but not on the inclination of FQ to the principal plane, except so far as that inclination affects  $\psi_1$ . The discovery of this circular symmetry about the axis  $F_2F$  came as a great

surprise. The points for which  $\Delta_1^*$  has any given value less than  $\Delta_1$  lie on an hyperboloid of revolution with foci  $F, F_2$ . In practice,  $\Delta_1/r$  is very small, of the order of  $10^{-4}$ . If we expand (3.9) in powers of  $\Delta_1/r$  as far as  $\Delta_1/r$ , we have  $r + \Delta_1^* = r + \Delta_1 \cos \psi_1$ . Hence, with sufficient accuracy,

$$\Delta_1^* = \Delta_1 \cos \psi_1. \quad (3.10)$$

To this approximation, the surfaces of constant  $\Delta_1^*$  are circular cones with  $F$  as vertex and  $F_2F$  as axis.

The results are easily extended to the case of multiple reflexions. Let  $F_3, F_4$  be the first and second images of  $F_2$ , and so the third and fourth images of  $F$ . The series can be continued. We have  $F_{2n}OF_{2n-2} = 2n\eta$ , and thus

$$F_{2n}OF = 2n\eta, \quad F_{2n}F = 2f \sin n\eta. \quad (3.11), (3.12)$$

With continued reflexions, the intensity of the reflected light diminishes so rapidly, even with semi-silvered surfaces, that we need not consider values of  $n$  exceeding 5, and thus  $n\eta$  is very small compared with a radian. We have

$$F_{2n}F = F_2F \sin n\eta / \sin \eta = nF_2F \left\{ 1 - \frac{1}{6}(n^2 - 1) \sin^2 \eta + \dots \right\}.$$

If  $\frac{1}{6}n(n^2 - 1) \sin^2 \eta \cdot F_2F < \frac{1}{10}\lambda$ , we have, as a good optical approximation,

$$F_{2n}F = nF_2F. \quad (3.13)$$

The straight line  $F_{2n}F$  is the  $n$ th cardinal axis of  $F$ .

A zigzag ray of order  $n$  reaches  $F$  from  $Q$  after reflexions at  $n$  points on  $U$  and at  $n$  points on  $V$ . The length of the zigzag is  $[QF]_n$ . The method used above gives

$$[QF]_n = QF_{2n}.$$

We then have  $\Delta_n^* = [QF]_n - QF = QF_{2n} - QF. \quad (3.14)$

We can show that  $QF_{2n} = Q_{2n}F$ .

If, in fig. 1, the luminous point be at  $R$  on the same side of  $U$  as  $F$ , the two rays which meet in  $F$  are  $EF$ , directed from  $Q$ , the image of  $R$  in  $U$ , and  $DF$ , where  $D$  is the point in which  $RF_1$  meets  $U$ . Each ray suffers one reflexion. The path of the ray suffering  $2n - 1$  reflexions is, after its first passage through  $U$ , identical with that of the ray from  $Q$ , directed to  $F_{2n}$ , after its first reflexion at  $U$ . The path difference is  $QF_{2n} - QF$ .

In the  $n$ th cardinal case, the path difference is  $\Delta_n$ , and, by (3.13),

$$\Delta_n = F_{2n}Q - FQ = F_{2n}F = n\Delta_1. \quad (3.15)$$

Let  $FQ$  make angle  $\psi_n$  with  $F_{2n}F$  produced. With

$$FQ = r, \quad F_{2n}Q = r + \Delta_n^*, \quad F_{2n}F = \Delta_n, \quad \text{and} \quad F_{2n}FQ = \pi - \psi_n,$$

we have

$$r + \Delta_n^* = \{r^2 + 2r\Delta_n \cos \psi_n + \Delta_n^2\}^{\frac{1}{2}}.$$

To the first power of  $\Delta_n$ ,

$$\Delta_n^* = \Delta_n \cos \psi_n = n\Delta_1 \cos \psi_n. \quad (3.16)$$

The angle  $\psi_n$ , which  $FQ$  makes with  $FF'_{2n}$ , the continuation of  $F_{2n}F$ , is not the same as the angle  $\psi_1$  between  $FQ$  and  $FF'_2$ . Now, by (3.11),  $F_{2n}OF_2 = 2(n - 1)\eta$ , and the  $F$ 's are on the circle of centre  $O$ . Hence

$F_{2n}FF_2 = (n-1)\eta$ . Any one of the three plane angles  $QFF'_n = \psi_n$ ,  $QFF'_2 = \psi_1$ , and  $F'_2FF'_{2n} = (n-1)\eta$ , which form a solid angle at F, is less than the sum of the other two. Hence  $\psi_n < \psi_1 + (n-1)\eta$ , and  $\psi_n > |\psi_1 - (n-1)\eta|$ . If  $\psi_1 = 0.05$ , and  $(n-1)\eta = 0.003$  radian,  $\psi_n$  lies between 0.053 and 0.047 radian, and  $\cos \psi_n$  between 0.99860 and 0.99890, and thus  $\cos \psi_n$  differs from  $\cos \psi_1$  (0.99875) by less than 0.016 per cent. When  $\eta$ ,  $\Delta_1$ ,  $\psi_1$ , and  $n$  are small enough, we have, as good optical approximations,

$$\Delta_n^* = \Delta_n \cos \psi_n = n \Delta_1 \cos \psi_1 = n \Delta_1^* \quad \dots \quad (3.17)$$

From Q there pass to F a direct ray and rays reflected 2, . . .  $2n$  times. If  $\Delta_1 = s\lambda$ , where  $s$  is integral, and if  $\Delta_1$  and  $\psi_1$ , . . .  $\psi_n$  be small enough, the path of the ray reflected  $2n$  times differs from that of the direct ray QF by  $ns\lambda$ , and then each reflected ray reinforces the direct ray. As already stated, no reversal of phase occurs at reflexion or at transmission in this ideal case. If  $\Delta_1 = (s + \frac{1}{2})\lambda$ , the ray reflected  $2n$  times reaches F with the same phase as the direct ray when  $n$  is even, and with opposite phase when  $n$  is odd. As a result, the contrast between "bright" and "dark" is softened.

#### § 4. Discussion of Aperture in Ideal Case.

If  $\Delta_1(1 - \cos \psi_1) < \frac{1}{4}\lambda$ , or  $\sin^2 \frac{1}{2}\psi_1 < \lambda/8\Delta_1$ , then  $\Delta_1 - \Delta_1^* < \frac{1}{4}\lambda$ , and, consequently, for no pair of rays reaching F from any point within the cone, vertex F, for which  $\sin^2 \frac{1}{2}\psi_1 < \lambda/8\Delta_1$ , will  $\Delta_1 - \Delta_1^*$  be as great as  $\frac{1}{4}\lambda$ . This indicates a great focusing effect. If  $\Delta_1 = 50\lambda$ , the condition is satisfied if  $\psi_1 < 5^\circ 43'$ , in striking contrast to the angle—perhaps  $2'$ —between the primal pair of rays. We can divide space by a series of cones, vertex F and axis  $F_2F$ , going by equal steps in  $\cos \psi_1$ , such that, if any point in one partition give a pair of rays which reinforce each other at F, any point in the next partition will give a pair which lessen each others effect. Some interesting conclusions can be drawn from this result, but they are not of practical importance, because the ideal case cannot be realized.

#### § 5. General Discussion on Aperture.

The rays do not come from a single point, but from an extended source. If the fringes which would be formed by a wedge of air or liquid, without flanking plates, lack vigour and definiteness of position, no optical appliance brought to bear on them could produce those qualities.

Let a single ray  $\Gamma$  in a principal plane  $\Pi$  fall on the simple wedge. On reaching the wedge, it splits into two rays which eventually pass through a single point F in  $\Pi$ , viz. a transmitted ray T and a twice reflected ray R. To every  $\Gamma$  in  $\Pi$  there is an F, and hence to every F there is a  $\Gamma$ . In the ideal case of an air wedge, we can either, as in § 3, find  $\Gamma$  when we are given F, or, as in § 10, can find F when we are given  $\Gamma$ . When the wedge has index  $\mu$ , there are exact but complicated formulæ by which F can be found when  $\Gamma$  is given; to find  $\Gamma$  exactly when F is given might be difficult.



Let the optical distance from a point Q on  $\Gamma$  to F by ray R exceed that by ray T by  $\Delta_1$ ; then  $\Delta_1$  is independent of the position of Q on  $\Gamma$ .

From every point of the source two rays, T\* (transmitted) and R\* (twice reflected) pass through the wedge, of index  $\mu$ , and, "really" or "virtually," reach F. If the optical path† of R\* exceed that of T\* by  $\Delta_1^*$ , and if  $\Delta_1 - \Delta_1^* < \delta$ , where  $0 < \delta < \frac{1}{4}\lambda$ , R\* and T\* will reinforce each other, if  $\Delta_1 = s\lambda$ , where  $s$  is an integer. The points for which  $\Delta_1^* = \Delta_1 - \frac{1}{4}\lambda$  lie on a "conical" surface; for all points within this surface,  $\Delta_1 - \Delta_1^* < \frac{1}{4}\lambda$ . A mental picture is easily obtained.

Let a spherical wave  $S_1$  start out from F at  $t=0$  and be reflected twice at the wedge. Let a second wave  $S_2$  start from F at  $t=\Delta_1/v$ , and be transmitted; here  $v$  is the velocity in air. The two wave-fronts reach a point Q on  $\Gamma$  at the same time and are then  $W_1, W_2$ . The fronts have  $\Gamma$  as a common normal and touch at Q. If a wave  $S_3$  starting from F at  $t=(\Delta_1-\delta)/v$ , where  $0 < \delta < \frac{1}{4}\lambda$ , be transmitted and become the wave-front  $W_3$ , it will cut  $W_1$  in a closed curve, and a volume  $\Omega$  will be enclosed by  $W_1$  and  $W_3$ . If Q\* be within  $\Omega$ , the optical distance from Q\* to F for R\* exceeds that for T\* by  $\Delta_1^*$ , and  $\Delta_1 - \Delta_1^* < \frac{1}{4}\lambda$ . If Q move off to infinity along  $\Gamma$ , the curve of intersection of  $W_1$  and  $W_3$  will trace out a "conical" surface such that, for every point within it,  $\Delta_1 - \Delta_1^* < \frac{1}{4}\lambda$ . Unless  $\mu=1$ , neither  $W_1$  nor  $W_3$  will be spherical, and the curve will not be a circle.

If we now put a flanking plate  $G_1$  in contact with the wedge and between it and the source, the three waves sent out from F will advance into the air beyond  $G_1$ . For every point within the surface marked out by the  $W_1$ - $W_3$  curve as it advances,  $\Delta_1 - \Delta_1^* < \frac{1}{4}\lambda$ .

The T\*-rays reaching F from points on the  $W_1$ - $W_3$  curve form a T\*-cone. The smaller  $\Delta_1/\lambda$  is, the greater is the solid angle of this cone. An exact estimate of the solid angle is not possible except for the ideal air wedge of §3. With  $\psi_1$  as in §3, the solid angle in the ideal case is  $\pi\lambda/2\Delta_1$ , and this may serve as a guide in the general case.

The vital point is that, to a cone of first order angle there corresponds a path discrepancy of the second order. The secret of the focusing power of the wedge is here.

The effect of multiple reflexions must be noticed. In the ideal case, we can show, by (3.16), that, if  $\Delta_n - \Delta_n^* < \lambda/4$ , then  $\Delta_1 - \Delta_1^* < \lambda/4n$ . If  $\Delta_1 = s\lambda$ , where  $s$  is an integer, we have  $\sin^2 \frac{1}{2}\psi_1 < 1/8ns$ . For any point within the cone defined by  $\psi_1 = 2 \sin^{-1}(8ns)^{-\frac{1}{2}}$ , those rays which suffer not more than  $2n$  reflexions will, on reaching F, differ in phase from the direct ray at F by less than  $\frac{1}{2}\pi$ . If rays from a larger cone reach F, or if reflexions above  $2n$  be of significance, confusion will grow at F. But it will be unorganized confusion, which merely weakens the contrasts by adding—it cannot subtract—illumination to bright and dark fringes

---

† If a reversal of phase occur at one reflexion, we must account for it in estimating the phase relation of R\* to T\* at F. See §6.



We can calculate  $\Delta$ . The phase change on reflexion at A or at B is either nil or a reversal. Let  $b$  be the ratio of the electric force in the wave along CE to that along AD, when these rays are due to GA, and  $a$  the ratio when they are due to KA; a negative value of  $a$  or  $b$  means a reversal of phase. Then,  $a, b$  have different values as the electric force  $E$  is (1) in, or is (2) normal to the principal section†. But  $a$  is always negative and  $b$  always positive‡. Hence, in either (1) or (2), when  $\Delta = n\lambda$ , where  $\lambda$  is the wave-length in air and  $n$  is integral, GA gives a bright and KA a dark fringe at F. Any wave, with GA as a ray, can be replaced by a (1)-wave and a (2)-wave. Each makes F bright. Hence F is bright for any direction of  $E$  in the incident wave. Similarly KA always makes F dark when  $\Delta = n\lambda$ . These results are independent of  $\mu$  and  $\mu'$  and of the angle of incidence. If  $a$  or  $b$  be small, as may happen in case (1), the effect at F is practically the same as if the ray EF did not exist.

### § 7. Sharpness of Fringes.

If, for each position of F on its locus OF, only the primal rays, DF, EF in fig. 3, reached F, the intensity would be of the form

$$r^2 + 2rs \cos(2\pi\Delta/\lambda + \phi) + s^2,$$

where  $\phi$  is a phase change independent of  $\Delta$ , the path difference for the two rays§. We could describe the pattern along OF either as bright fringes on a less bright field, or as dark fringes on a brighter field. The wave-front W, starting from any point Q in a principal plane, can be split into  $W_1$  and  $W_2$ , coincident in position with W. In  $W_1$ , the electric force at a point in that plane is in that plane; in  $W_2$ , it is normal to that plane. If the primal  $W_1$ -rays make F bright and F' dark, the multiply reflected  $W_1$ -rays have a limited effect and leave F bright and F' dark, but they change the distribution of I, the intensity. When the fringes are observed with transmitted light, the brightness is more concentrated than the darkness. Near F, the centre of a bright band, I diminishes rapidly as the distance from F increases, but near F', the centre of a dark band, I increases slowly with the distance from F'. The pattern is one of narrow bright bands on a dark field. In the case of reflexion the conditions are reversed, and narrow dark bands are seen on a bright field. The sharpening effect increases with the reflecting power of the two surfaces bounding the wedge. The primal  $W_2$ -rays make the same points F, F' bright and dark, as do the primal  $W_1$ -rays.

† G. F. C. Searle, 'Experimental Optics,' § 308.

‡ The statement that  $a$  is always positive and  $b$  always negative is true when the angle of the wedge is exactly zero. Actually the wedge has small angle  $\eta$ . If AB make  $\alpha, \beta$  with the normals at A, B, and if  $\gamma$  be the polarizing value of  $|a|$  or of  $|\beta|$ , a small range of values of  $|a|$  exists within which  $|a| - \gamma$  and  $|\beta| - \gamma$  have opposite signs. In this range  $a$  is positive and small, and  $b$  is negative and very small.

§ A complete theory would take account of the differences of directions of the rays concerned. But those differences are so small that, in a working approximation, they may be neglected.



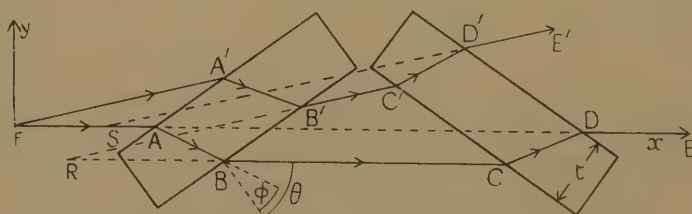
Let the electric force, at the interface, in the incident wave corresponding to AB (fig. 3) be  $E_0$ , and in the reflected wave corresponding to BC be  $E_R$ , and let  $E_R/E_0=c$ . Then  $c$  may be positive or negative, and  $c^2 < 1$ . The greater  $c^2$  is, the narrower are the bright transmission bands and the dark reflexion bands. In each system the contrast increases with  $c^2$ . This applies to both the  $W_1$ - and the  $W_2$ -bands. But  $c^2$  is different in the two cases. If the light be so polarized that  $E$  is in the principal plane, only the  $W_1$ -wave exists, and now the angle of incidence can be so chosen that  $c^2$  is small, and then the bands will be broad and the contrast weak. In the limit, at the polarizing angle, the field by transmission will be all bright and that by reflexion all dark. There is now no reflexion at any surface of the flanking plates.

### § 8. Optical Effect of Glass Plates.

In the experiments, two glass plates  $G_1$ ,  $G_2$  flank the wedge, and  $G_2$  lies between the wedge and the observer. We are not concerned with the effect of  $G_1$ , but we must consider that of  $G_2$ . The rays reach the wedge through  $G_1$ , when transmitted, or through  $G_2$  when reflected. In tracing the rays after they leave the wedge, we may consider  $G_2$  as infinitely near the face of the ideal wedge.

Let the horizontal plane FAB (fig. 4) be a principal plane of the wedge. Then the faces of  $G_2$  are vertical. Let rays from a real or virtual object point F meet  $G_2$ , which has thickness  $t$  and index  $\mu$ . Let the chief ray FA make  $\theta$ , and the refracted ray AB make  $\phi$ , with the normal. We take F as origin, FA as axis of  $x$  and Fy as in fig. 4. The emergent ray BC is parallel to FA. Let the abscissæ of A, B... be  $a$ ,  $b$ ... Then

Fig. 4.



$b-a=t \cos (\theta-\phi) \sec \phi$ . The emergent pencil has focal lines,  $L_1$  vertical,  $L_2$  horizontal, which intersect CB produced in  $J_1$ ,  $J_2$ . Then  $j_1=b-a-(t/\mu) \cos ^2 \theta \sec ^3 \phi$ , and  $j_2=b-a-(t/\mu) \sec \phi$ . The line BC is  $y=-t \sin (\theta-\phi) \sec \phi$ .

In the apparatus of § 17, a vertical compensating plate  $G_3$ , equal to  $G_2$ , is used; it is between  $G_2$  and the observer, and its faces make  $2\theta$  with those of  $G_2$ . Hence its normals make  $\theta$  with BC. The ray FABC meets  $G_3$  at C and emerges from it at D along DE, in line with FA. The emergent pencil has focal lines,  $L_3$  vertical,  $L_4$  horizontal, which meet FE in  $J_3$ ,  $J_4$ . Then  $j_3=2j_1$ ,  $j_4=2j_2$ .



Let a ray  $FA'$  in the plane  $FAB$  make a small angle  $\gamma$  with  $FA$ , and let its course be  $FA'B'C'D'E'$ . Let  $\gamma$  be positive when the angle of incidence at  $A'$  exceeds  $\theta$ . Then  $B'C'$  meets  $BC$  in  $R$ , and  $D'E'$  meets  $FAD$  in  $S$ . As far as  $\gamma^2$ ,

$$r = t \cos(\theta - \phi) \sec \phi - (t/\mu) \cos^2 \theta \sec^3 \phi \{1 - m\gamma - n\gamma^2\},$$

and

$$s = 2t \cos(\theta - \phi) \sec \phi - 2(t/\mu) \cos^2 \theta \sec^3 \phi (1 - n\gamma^2),$$

where, with  $q = \cos \theta / \mu \cos \phi$ , we have  $m = \frac{3}{2} \tan \theta (1 - q^2)$ , and

$$n = \frac{1}{2} (1 - q^2) \{1 - \tan^2 \theta (1 - 5q^2)\}.$$

Let a ray  $FA''$ , whose projection on  $FAB$  is  $FA$ , make  $\alpha$  with that plane. Then the equation of the projection of  $B''C''$  is

$$y = -t \sin(\theta - \phi) \sec \phi - \frac{1}{2} \alpha^2 t \Omega \sin \theta,$$

and  $B''C''$  meets its own projection at  $U$ , where

$$u = t \{ \cos(\theta - \phi) - 1/\mu \} \sec \phi - \frac{1}{2} \alpha^2 t \Omega \cos \theta.$$

Here

$$\Omega = (\mu^2 - 1) \cos \theta \sec^3 \phi / \mu^3.$$

The ray  $FA''$  gives rise to the ray  $D''E''$  which emerges from the plate  $G_3$ . The projection of  $D''E''$  coincides with  $FA$ , and  $D''E''$  meets  $FA$  in  $V$ , where  $v = 2u$ .

The plate  $G_3$  is of great advantage in the experiment of § 17, as it corrects the lateral displacement, due to  $G_2$ , of the fringes. It also makes the aberration of rays in the principal plane symmetrical and proportional to  $\gamma^2$ . When  $G_3$  is not used, the aberration is one-sided, with its leading term proportional to  $\gamma$ . The lessening of the aberration improves the crispness of the pattern and diminishes the uncertainty in locating the (vertical) fringes.

### § 9. Fringes in Real Life.

Fringes are judged to exist when a series of bands, alternately "bright" and "dark," are seen. The variations of intensity may impress themselves on the eye, either directly or through a microscope. If the fringes be photographed, the contrasts are examined by eye or perhaps by a micro-photometer. In any case, the differences of intensity needed to convince the observer that fringes exist are small compared with the maxima of intensity. If the observer look through a "wedge" of non-silvered glass plates towards a sodium flame, fringes distinct enough for measurement are seen, although the differences between the maxima and minima of intensity are only a few per cent. The fringes are visible against the full glare of the flame.

A better method, and more under control, is to use semi-silvered plates, as in the experiment of § 17. The observer, using a microscope or the naked eye, looks at the fringes due to a sodium flame on the other side of the wedge. A second sodium flame is placed so that its light is reflected into the microscope or the eye by a sloped plate such as  $R$  (fig. 10). The contrasts are surprisingly little weakened by the light from the second source. The fringes due to a sodium flame are still



Thus in fig. 5,

$$AF = LF - LA = OA \sin \theta - LA = OA \sin \theta - AN \cos \theta. \quad (10.1)$$

With fig. 6,

$$AF = OA \sin \theta + AN \cos \theta.$$

In the experiment of § 17,  $LA < 0.005$  cm. If we neglect  $LA$  in comparison with  $OA \sin \theta$ ,

$$AF = OA \sin \theta. \quad (10.2)$$

To this approximation,  $F$  is on the circle of diameter  $OA$ .

In one method of finding  $\Delta_1$ , we use  $F_1, F_2$ , where  $F_1$  is the image of  $F$  in  $V$ , and  $F_2$  is that of  $F_1$  in  $U$ . Then  $F, F_1, F_2$  are on a circle of centre  $O$ . Now  $FF_1, F_2F_1$  are normal to  $V, U$ , and hence  $FF_1F_2 = \eta$ , and  $FOF_2$ , at  $O$ , is  $2\eta$ . If  $M$  be the mid-point of  $FF_2$ ,  $MOF = \eta$ , and  $OMF = \frac{1}{2}\pi$ . A ray  $F_1A$ , reflected at  $U$ , becomes  $AQ$ , and hence  $QA$  is directed to the image of  $F_1$  in  $U$ , *i. e.* to  $F_2$ ; thus  $QAFF_2$  is a straight line. Now  $BF = BF_1$ , and  $AF_2 = AF_1$ , and hence, in fig. 5,

$$\Delta_1 = AB + BF - AF = AF_1 - AF = AF_2 - AF = FF_2. \quad (10.3)$$

Thus

$$\Delta_1 = 2OF \sin \eta = 2ON \cos \theta \sin \eta = 2AN \cos \theta. \quad (10.4)$$

With regard to fig. 6, see (3.5).

The system of  $F$ 's is useless when the index of the wedge is other than unity. For comparison with the analysis in that case, we find  $\Delta_1$  without using  $F_1$  or  $F_2$ . Since  $AB = KB$ , we have  $\Delta_1 = KF - AF$ . Now  $KAN = \eta$ , and thus, in fig. 5,  $KAQ = \theta + \eta$ . But  $KFA = 2\eta$ , and hence  $FKA = \theta - \eta$ . With  $AK = 2OA \sin \eta$ , and  $AN = OA \tan \eta$ , we have

$$AF = AK \sin (\theta - \eta) / \sin 2\eta = OA \sin (\theta - \eta) / \cos \eta = OA \sin \theta - AN \cos \theta, \quad (10.5)$$

and

$$AL = AN \cos \theta = OA \tan \eta \cos \theta = OM \tan \eta = MF. \quad (10.6)$$

We have

$$KF = AK \sin (\theta + \eta) / \sin 2\eta = OA \sin (\theta + \eta) / \cos \eta.$$

Hence

$$\Delta_1 = OA \{ \sin (\theta + \eta) - \sin (\theta - \eta) \} / \cos \eta = 2OA \tan \eta \cos \theta = 2AN \cos \theta. \quad (10.7)$$

Let  $OM$  be the perpendicular from  $O$  on  $AF$ ; then  $AOM = \theta$ , and  $OM = OA \cos \theta$ . If  $FOM = \omega$ ,

$$\tan \omega = (AM - AF) / OM = (OA \sin \theta - AF) / OM.$$

Then, by (10.5), (10.6),

$$\tan \omega = AN \cos \theta / OM = \tan \eta,$$

and

$$\omega = \eta. \quad (10.8)$$

In the case of fig. 6, where  $\theta$  is negative, slight changes are needed, but (10.7), (10.8) hold good.

§ 11. *Formation of Fringes.*

So far we have considered a single point F. But, within the limits imposed by the apparatus, the conditions are identical at all points on a line through F perpendicular to a principal plane. Hence, if there be a maximum of illumination at F, there will be a bright fringe through F perpendicular to the principal plane.

§ 12. *Spacing of Fringes along Chord OF.*

The chord OF (figs. 5, 6) makes angle  $\theta$  with Ov, for O, F, A, N are on a circle, and  $FAN = \pi - \theta$  in fig. 5, and  $FAN = \theta$  in fig. 6. An observer, looking through the wedge towards a sodium flame, sees a series of fringes. The flame gives, among others, a set of rays in the principal plane of F and parallel to FA in figs. 5, 6. These rays make  $\theta$  with a normal to U. For each position of A on Ou, there is a corresponding point F on the fixed line OG, which makes  $\theta$  with Ov, and  $\eta$  with OM. For any point F on OG, the path difference for the rays QABF, QAF is  $\Delta_1 = 2OF \sin \eta$ . If  $\Delta_1 = s\lambda$ , where  $s$  is integral, and if there be no change of phase, the two rays reinforce each other. The position of F as a maximum of illumination is determined by this primal pair of rays, and so a bright fringe passes through F. Thus, at a bright fringe,

$$OF = \frac{1}{2}s\lambda / \sin \eta.$$

If there be  $q$  fringe spaces per unit length of OG,  $q = s/OF$ , and thus

$$q = 2 \sin \eta / \lambda. \quad \dots \dots \dots (12.1)$$

Thus  $q$  does not involve  $\theta$  and so is independent of the direction of OG, and therefore of that of FA—a result of great practical importance. When  $\lambda$  is known, we can find  $\eta$ , for

$$\sin \eta = \frac{1}{2}q\lambda. \quad \dots \dots \dots (12.2)$$

The distance  $1/q$  from fringe to fringe is  $\frac{1}{2}\lambda / \sin \eta$ —the fundamental fringe spacing.

The spacing of the fringes is measured against the eye-piece scale of a microscope whose axis is along FA. For a given  $\theta$ , the fringes do not lie along a line perpendicular to the axis FA, but lie along OG, which makes  $\eta$  with that perpendicular. The angle  $\eta$  is very small; its effect on the observed spacing is negligible.

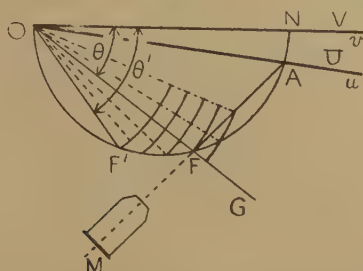
§ 13. *Distribution of Fringes on Fringe Circle ONA.*

Let a series of O-circles, with O as centre and radii increasing from zero by steps of  $\frac{1}{2}\lambda / \sin \eta$ , be drawn in the plane ONA, as in fig. 7. Let OG make angle  $\theta$  with Ov. Then the “bright” fringes, due to rays which are parallel to the plane ONA and have  $\theta$  as angle of incidence on U, are normal to ONA, and pass through the points in which OG intersects the O-circles. If a change of phase occur, “dark” replaces “bright.” But each bright (or dark) fringe, corresponding to a given A, meets the corresponding fringe circle ONA, and hence the bright fringes associated



with A pass through the points in which the fringe circle ONA intersects the O-circles. Dark A-fringes occur where the fringe circle is cut by O-circles of radii  $\frac{1}{2}(s+\frac{1}{2})\lambda/\sin \eta$ , where  $s$  is integral.

Fig. 7.



Let FA turn in the principal plane about A, and let Q move so as to keep on FA. Let F, F', two points on the circle ONA, be such that, with Q on FA, there is a bright fringe at F, and, with Q on F'A, there is one at F'. Let  $\text{FON}=\theta$ ,  $\text{F'ON}=\theta'$ . By § 10,  $\theta$ ,  $\theta'$  are the angles of incidence of the two rays QA, Q'A at A. Now ON is a diameter of ONA, and hence  $\text{OF}=\text{ON} \cos \theta$ ,  $\text{OF}'=\text{ON} \cos \theta'$ . Let  $\text{OF}-\text{OF}'$  equal  $m$  fringe intervals, each of the fundamental length  $\frac{1}{2}\lambda/\sin \eta$ . Then

$$\text{OF}-\text{OF}'=\text{ON}(\cos \theta-\cos \theta')=\frac{1}{2}m\lambda/\sin \eta. \quad (13.1)$$

But  $\text{ON} \sin \eta=\text{AN}$ , and thus

$$\text{AN}=\frac{1}{2}m\lambda/(\cos \theta-\cos \theta'). \quad (13.2)$$

If a microscope M, with its axis directed to A, were to move round with AF,  $m$  fringe intervals would appear to pass its cross-wire, while FON changed from  $\theta$  to  $\theta'$ . To keep the fringes in focus, it would be necessary to move M along its axis MA.

#### § 14. Light Sources.

The theory assumes the light to have a single wave-length. If we use a mercury vapour lamp, we can isolate the green line,  $\lambda=5.4607 \times 10^{-5}$  cm. by a filter. With a sodium flame, there are two lines  $D_1$ ,  $D_2$ , and  $\lambda_1=5.896$ ,  $\lambda_2=5.890 \times 10^{-5}$  cm. If the distance AN at the centre of the wedge be only a few wave-lengths, the fringes due to  $D_2$  will nearly coincide with those due to  $D_1$ . If  $\text{AN}=100 \lambda$ , and  $\eta=5 \times 10^{-4}$  radians, the  $D_1$  and  $D_2$  fringes corresponding to A will, with  $\theta=0$ , be out of step by about 0.01 cm.; the fringe spacing is about 0.0589 cm. The observer's estimate of the spacing will lean towards one corresponding to a wave-length between the mean  $\lambda=5.893 \times 10^{-5}$  cm. and  $\lambda_2$ , because the  $D_2$  is more intense than the  $D_1$  line.

A neon lamp is convenient. There is a strong yellow line of  $\lambda=5.8525 \times 10^{-5}$  cm., and many red lines and others of much less intensity. The patterns due to the red lines so overlap as to produce a nearly uniform background to the yellow fringes.



the plates and the observer, the thin end of the wedge is towards him. In this case, a horizontal pin may be placed so that there is no parallax between it and a fringe. If, with incidence at  $45^\circ$ , the fringes be 5 to 10 cm. from the centre of the plates, the artist may be satisfied. Plates so assembled will keep stable for a long time.

The plates are mounted, with the strip T vertical, on a holder J (fig. 8), standing on a sheet of paper on a board. The thin end of the wedge projects from J, so that A, its centre, is clear of J. With a sodium flame at S, the fringe by reflexion, which corresponds to A, will be found at F; the fringes can be "caught" on ground glass. If the observer, his eye near E, stick a pin into the board so that it has no parallax against the A fringe, the pin marks F. If the flame be moved to (S), the A-fringe will be at (F) behind the wedge, and is viewed from (E). The pin will be lost in the glare of the reflected light, but, if a source of small intensity, *e. g.* a neon lamp, be placed in the line A(F), as shown by the small circle, the pin is easily seen. The thickness of the flanking plates prevents the locus of F from being exactly a circle on OA as diameter.

The fringes can be seen also by transmission. The A-fringe will be at F or at [F] according as the source is at T or at [T]. If, with the source at S, there be a bright A-fringe at F, there will be a dark A-fringe at the same point F, when the source is at T (see § 6).

A thin vertical rod, pointed at both ends, and sliding vertically in a block is more convenient than a pin. When the fringe has been located, the rod is pressed down, against a spring, to record the position on the paper.

#### § 16. *Second Experiment. Measurement of Angle $\eta$ of Air Wedge.*

In the usual experiment, the axis of the microscope used in measuring the fringe spacing is approximately normal to the plates flanking the wedge, and the incidence of the rays ultimately reaching the observer is nearly normal. With non-silvered plates, the pattern seen by reflected light is "soft," and may be described as diffuse dark bands on a bright ground or diffuse bright bands on a dark ground. A great improvement is easily obtained; the method depends on (12.1).

The plates may be 2.5 cm. square, selected for relative flatness. They touch along one common edge; the opposite edges are separated by a strip of tin-foil 0.0015 cm. thick and 0.3 cm. wide. This will give about 20 for  $q$ , the number of fringes per cm. At the centre, the wedge will be about  $12\lambda$  thick. The plates are waxed together as in § 15. When a vertical travelling microscope is used, the plates are set so that (1) the normal to either makes about  $45^\circ$  with the vertical, and (2) the edge of the wedge is horizontal. The azimuth of the wedge is adjusted to make the fringes perpendicular to the (horizontal) track of the microscope. The chief ray of the pencil used by the microscope is vertical. By § 12, the fringes lie in a plane *through the edge*, and the normal to this plane makes  $\eta$  with the axis of the microscope. We may treat the plane as

horizontal. As the microscope moves along its track, the fringes remain in focus. The distance from fringe to fringe is  $q^{-1}$ , and  $q = 2 \sin \eta / \lambda$ ; the angle of incidence does not appear.

The pattern is of narrow dark fringes on a bright ground (see § 7). The contrasts are strengthened if the lower face of the lower plate be coated with black varnish, and are further intensified if a horizontal rod 0.3 cm. thick be held in a suitable position near the wedge to prevent unwanted light from reaching the eye-piece of the microscope. The best position is easily found. With good adjustments, the fringes are remarkably dark and remarkably narrow (see fig. 16).

### § 17. Apparatus for Third Experiment.

The air wedge is flanked by two pieces of plate glass, about 5 cm. square, of equal thickness 0.65 cm.; they are selected for flatness. One face of each plate is semi-silvered. The silvered faces are separated by a vertical band  $T_1$  (fig. 9), formed of three layers of tin-foil, and by a pad  $T_2$  of four layers. This will give, with sodium light, fringes spaced at about

Fig. 9.

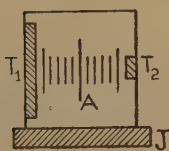


Fig. 10.

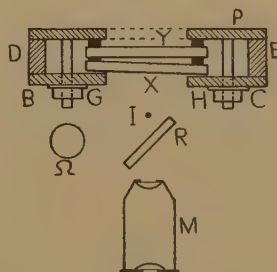
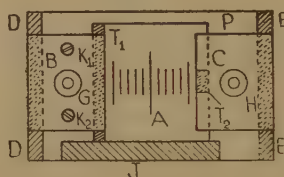


Fig. 11.



$\frac{1}{2}$  mm. The distance of the "edge" Oz (§ 3) of the wedge from the centre of the plates will be about 15 cm.; it depends somewhat on the extent to which the layers of foil are compacted. With careful work, the fringes will be nearly parallel to  $T_1$ . A slight adjustment of their direction can be made by the screws  $K_1$ ,  $K_2$  described below. A scale of 20 mm., with its lines parallel to  $T_1$ , is ruled on one of the semi-silvered faces. The long central line marks the point A of figs. 5, 6.

An optical bench has carriages for the wedge and the microscope. A horizontal table turns about a vertical shaft rising from the wedge carriage. A circular scale turning with the table is read by a pointer fixed to the carriage. The vertical metal plate P (figs. 10, 11), which carries the wedge plates, is fixed to the table by means of a metal strip of L-section. Two slots in the horizontal part of the L, cut at right angles to the length of the strip, accommodate two screws, and allow the corresponding adjustment of the strip on the table. The plate P is held against the vertical part of the L by screws passing through two horizontal slots in P. This arrangement allows any point of the wedge, within limits, to be brought to the axis of the shaft.



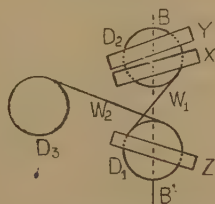
The wedge plates X, Y (fig. 10) are held against P by the clamps B, C, and rest on the horizontal ledge J. The ledge helps when the plates, with their delicate films of silver, are being put into position. One end of B rests on a block D, and one end of C on E. Pressure is applied by nuts G, H, with spring washers, screwing on studs fixed to P. The plate Y rests on three projections, two at the B-end and one at the C-end. Clamp B is shaped to touch X at two points corresponding to the two B-projections, and C touches X at one point opposite the C-projection. The blocks are shaped so that B touches D at a central point, and C touches E at two points. Two screws  $K_1$ ,  $K_2$  (fig. 11), working in B, can be made to press against the plate P. A small pressure exerted by  $K_1$  or  $K_2$  against B sometimes helps in persuading the fringes to be vertical. Once the system has settled down, which may take a little time, it is generally very stable.

A slow angular motion can be given to the wedge by a screw turning in a fitting attached to the wedge carriage. The free end of an arm, which can be clamped to the sleeve supporting the table, is drawn against the screw by a spring. Two stops on the carriage, one fixed and the other adjustable, limit the movement of the arm.

A microscope M (fig. 10) is used to locate the fringes; objects 14 cm. from the objective are seen in focus. Its eye-piece scale of  $\frac{1}{2}$  mm. was formed by cutting nicks, on a dividing engine, in the sharp edge of a thin plate of brass. A central nick S defines the line of collimation. The scale is used in measuring the fringe spacing when the fringes are at a distance from the wedge. It may be calibrated by aid of the scale ruled on the silver.

A plate Z (fig. 12) of glass of the same quality and thickness as the

Fig. 12.



plates X, Y is mounted on a vertical shaft near the wedge. The plane containing the axes of this shaft and of the table is parallel to the axis of the bench. The shaft carries a drum  $D_1$ , and the sleeve bearing the table carries an equal drum  $D_2$ . A thin wire  $W_1$ , fixed to  $D_1$  and to  $D_2$ , is wound two or three times round each. A second wire  $W_2$ , fixed to  $D_1$ , makes two or three turns round  $D_1$ , and then passes to a drum  $D_3$  turning on a vertical shaft. A helical spring inside  $D_3$ , with one end fixed to its shaft and the other to  $D_3$ , gives a suitable tension to  $W_2$ , and, through  $W_2$ , to  $W_1$ . The drums  $D_1$ ,  $D_2$  turn in opposite directions. The holder carrying Z is, removably, located in a fitting adjustable relative to  $D_1$ . If Z be parallel to X when X is perpendicular to the axis of the bench,

then X and Z always make equal angles, in opposite directions, with that axis. After Z has been adjusted, the microscope, with its axis parallel to that of the bench, is set so that the mark A on the silver is in focus on the centre S of the eye-piece scale. Then, however the wedge be turned, the image of A can always be brought to S by sliding the microscope along the bench. If it be focused first on A and then on the fringes, the difference of bench readings gives the distance which would be found between A and the fringe corresponding to A, if X were infinitely thin and Z were absent.

### § 18. *Experimental Details.*

The assembly of the apparatus involves the adjustment of the wedge on the table. The table is set so that the plates X, Y are, as nearly as can be judged by eye, in Position 1, in which they are perpendicular to the bench axis and X is the nearer to the observer's end of the bench. A carriage bearing a suitable object is adjusted so that a distance rod touches the object and X; the axis of the rod is parallel to the bench and intersects the axis of the table shaft. The table is then turned through  $180^\circ$  to Position 2, and the distance between Y and the object is tested by the rod. Any divergence is halved by moving the L-strip on the table at right angles to its length. Then the microscope, its axis parallel to the bench, is focused on the mark A when the wedge is in Position 1. The wedge is turned to Position 2, and any divergence is halved by sliding the plate P relative to the L-strip.

Before the observations are made, the zero reading of the circular scale, viz. that corresponding to Position 1, is found. The plate Z is removed. A vertical pin I (fig. 10) is set about 3 cm. from the plate X of the wedge, light from a lamp  $\Omega$  is directed past it by a glass plate R, between I and the microscope M, and M is adjusted so that I is in focus on the scale-centre S. For convenience, R and I are mounted on a plate which can be fixed to a stand. The M-carriage is then set so that the image of I at the semi-silvered surface of X is in focus, and the table is turned by the slow motion screw to bring the image of I to S. The wedge is now in Position 1, and the circle reading is taken. The table is clamped in this position. The plate Z is now mounted on its shaft, and by R and I, now between Z and M, Z is set perpendicular to the bench; the adjustment is made by moving, relative to drum  $D_1$ , the fitting which locates Z. A screen is placed between Z and X during this process.

The semi-silvering makes the bright fringes narrow compared with their common interval. But they have no sharp edges, and it is difficult to decide when they are best in focus. An improvement is gained in the following manner. A collimator, of focal length 10 cm., is placed between the source and the wedge, with its lens towards the wedge. A plate, with a vertical slot 0.3 cm. wide, slides in a horizontal groove, and is in the focal plane of the lens. The observer at the microscope

now sees fringes only in a patch corresponding to the slot. If the plate be moved to and fro in its groove, the fringes in the patch move relative to the eye-piece scale unless they be in focus. This result can be tested by focusing on the mark A, which gives a sharp image. If, as seen in the microscope, the fringes move in the same direction as the plate, the microscope must be moved nearer to the wedge.

The observations can now be made. The wedge circle is set to its zero; then it is in Position 1. and  $\theta=0$ . The microscope is focused on the scale on the silvering, and the number  $q$  of fringe spaces in one cm. is found. This may be repeated for Position 2, with  $\theta=180^\circ$ . By (12.2),  $\sin \eta = \frac{1}{2}q\lambda$ . The number  $w$  of fringe spaces in the length of the eye-piece scale is also found in the two Positions.

The locus of the F corresponding to A is now explored. The wedge is set at  $\theta=15^\circ, 30^\circ, 45^\circ$ , and, for each angle, the bench readings of the microscope when in focus (1) on A and (2) on the fringes, *i. e.* on F, are taken; the difference is FA. The number  $w$  of fringe spaces in the length of the eye-piece scale is observed for each angle. By the theory (§ 10),  $FA=OA \sin \theta$ , with sufficient accuracy, and, by § 12,  $w$  is constant.

The observations are repeated for  $\theta=-15^\circ, -30^\circ, -45^\circ$ , and also for  $\theta=\pm 15^\circ, \pm 30^\circ, \pm 45^\circ$  from Position 2. The mean value of  $|FA/\sin \theta|$  is taken as OA. Since we know  $\sin \eta$ , we can now find AN, for  $AN=OA \tan \eta$ .

An independent determination of AN can be made. The table is set at  $\theta=0$  (Position 1), and is clamped with the slow motion arm against the fixed stop. The screw is then turned to bring the arm against the adjustable stop, which may be set so that the table moves from  $0^\circ$  to about  $25^\circ$ . The microscope is gradually moved along the bench to keep the fringes in focus, as the table is turned, and the number  $m$  of fringe spaces crossing the centre S of the eye-piece scale is noted. The stops allow the readings to be repeated. Other observations may be made in which  $\theta$  is varied from about  $25^\circ$  to  $35^\circ$ , and from about  $35^\circ$  to about  $45^\circ$ , and also from  $0^\circ$  to  $-25^\circ$ ,  $-25^\circ$  to  $-35^\circ$ , and  $-35^\circ$  to  $-45^\circ$ . The table is then set in Position 2, and similar observations are made. The angles are chosen to make the number of fringe transits roughly the same in each range. The actual values of  $\theta$  for the beginning and end of each range must, of course, be recorded.

### § 19. Practical Example.

Sodium light, mean  $\lambda=5.893 \times 10^{-5}$  cm. was used.

I. *Determination of  $\eta$ .*—Number of fringe spaces to 1 cm. on the scale ruled on silvering happened to be 20.0. Hence, by (12.2)

$$\sin \eta = \frac{1}{2}q\lambda = \frac{1}{2} \times 20 \times 5.893 \times 10^{-5} = 5.893 \times 10^{-4}.$$

Thus

$$\eta = 5.893 \times 10^{-4} \text{ radians} = 2' 1.6''.$$

II. *Determination of OA.*

Circle reading.	$\theta$ .	Microscope readings.		FA .	Fringes to 10 div. of eye-piece scale = $w$ .
		Mark A.	Fringes.		
°	°	cm.	cm.	cm.	
187	0	66.88	66.88	0	23.5
172	15	66.90	65.07	1.83	23.5
157	30	66.99	63.43	3.56	23.5
142	45	67.17	61.98	5.19	23.5
202	—15	67.04	69.01	1.97	23.5
217	—30	67.03	70.73	3.70	23.5
232	—45	67.33	72.85	5.52	23.5
367	0				23.75
382	15	67.17	65.00	2.17	23.5
397	30	67.26	63.47	3.79	23.5
412	45	67.40	62.18	5.22	23.6
352	—15	67.20	69.00	1.80	23.5
337	—30	67.30	70.95	3.65	23.5
322	—45	67.42	72.64	5.22	22.3

Taking mean values, we have

$ \theta $ .	FA  obs.	$ \sin \theta $ .	FA/sin $\theta$   obs.	FA  cal.
°	cm.		cm.	cm.
15	1.942	0.2588	7.504	1.925
30	3.675	0.5000	7.350	3.720
45	5.288	0.7071	7.478	5.260

Sums                      10.905                      1.4659

With  $x=|\sin \theta|$ , and  $y=|FA|$ , the co-ordinates of the centroid of the three points are  $\bar{x}=10.905/3$ ,  $\bar{y}=1.4659/3$ . Hence  $\bar{x}/\bar{y}=7.439$ . By (10.2)

$$OA = FA/\sin \theta = 7.439 \text{ cm.}$$

The values of |FA| calculated, from  $FA=OA \sin \theta$ , with  $OA=7.439$  cm. are given in the last column

III. *Determination of AN.*—Zero reading of circle for Position 1 was 187°.

Circle Reading.		$ \theta $ .		$ \cos \theta - \cos \theta' $ .	m .	$\frac{m}{ \cos \theta - \cos \theta' }$ .
Initial.	Final.	Initial.	Final.			
°	°	°	°			
187.0	209.4	0.0	22.4	0.0755	12.0	158.9
209.0	232.0	22.0	45.0	0.2201	34.0	154.5
187.0	164.4	0.0	22.6	0.0768	11.75	153.0
367.0	344.3	0.0	22.7	0.0775	11.6	149.7
344.3	322.0	22.7	45.0	0.2154	32.0	148.6
367.0	389.8	0.0	22.8	0.0781	12.0	153.6
389.8	412.0	22.8	45.0	0.2148	33.5	156.0

Sum                      1074.3



Mean value  $|m|/|\cos \theta - \cos \theta'| = 1074 \cdot 3/7 = 153 \cdot 5$ .

By (13.2),

$$AN = \frac{1}{2} \lambda m / (\cos \theta - \cos \theta') = \frac{1}{2} \times 5 \cdot 893 \times 10^{-5} \times 153 \cdot 5 = 0 \cdot 004523 \text{ cm.}$$

From geometry of wedge,

$$AN = OA \tan \eta = 7 \cdot 439 \times 5 \cdot 893 \times 10^{-4} = 0 \cdot 004384 \text{ cm.}$$

I had hoped to give results in III, with angles recommended in § 18, but opportunity failed.

### § 20. Introduction to Fourth Experiment.

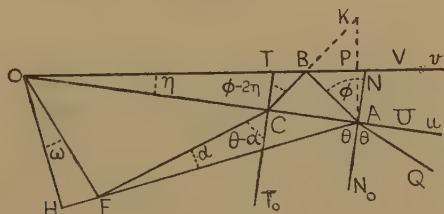
The refractive index of a liquid may be found by the Newton's rings formed by a plate and a lens with (1) air, and (2) liquid between them. With normal incidence, the rings for the liquid are very faint †.

In the present experiment, two plane plates inclined at a small angle are used, and the fringes are viewed with reflected light. With an angle of incidence of  $30^\circ$  or more, the fringes, when there is water between the plates, may be made vigorous enough for good measurement. The index of the liquid can be found from the angle of incidence and the ratio of the fringe spacing with air to that with liquid between the plates.

### § 21. Location and Spacing of Fringes formed by a Refracting Wedge.

The wedge is supposed to be flanked by two mathematical planes U, V, which reflect light, but do not themselves deviate any ray passing through them. The planes U, V (fig. 13) meet the plane of the diagram at right angles in Ou, Ov. The small angle uOv is  $\eta$ . The index of the wedge is  $\mu$ .

Fig. 13.



Let a ray QA in the principal plane uOv meet U at A. It gives rise to a reflected ray AF, and to a refracted ray AB; the latter, after reflexion at B on V and refraction at C on U, emerges along CF to meet AF in F. Let  $N_0AN$  and  $T_0CT$  be normal to U. Let QA, AB make  $\theta$ ,  $\phi$  with  $N_0AN$ ; let  $\theta$  be positive when  $OAQ > \frac{1}{2}\pi$ . Let CF, BC make  $\theta - \alpha$ ,  $\phi - \beta$  with  $T_0CT$ ; then  $AFC = \alpha$ . If V were parallel to U,  $\beta$  would vanish. The turning of V through  $\eta$  turns the reflected ray through  $2\eta$ , and thus  $\beta = 2\eta$ . Exact formulæ are easily found but are unserviceable by complexity; approximations, valid when  $\eta$  is very small, are used.

† In the ordinary method, the glare due to unwanted reflexions nearly swamps the ring pattern for the liquid. In a method recently developed by Mr. R. L. Woolley and me, unwanted reflexions are eliminated, and vigorous rings are seen for water and with nearly normal incidence.

We first find  $\alpha/\eta$ . With  $\sin \theta = \mu \sin \phi$ , and  $\sin (\theta - \alpha) = \mu \sin (\phi - 2\eta)$ , subtraction gives

$$\sin \frac{1}{2}\alpha = \mu \sin \eta \cos (\phi - \eta) / \cos (\theta - \frac{1}{2}\alpha). \quad (21.1)$$

When  $\eta$ , and so  $\alpha$ , are very small, we have

$$\alpha = 2\mu\eta \cos \phi / \cos \theta. \quad (21.2)$$

If  $\theta = \frac{1}{2}\pi$ , (21.1) fails as an approximation. When  $\theta = \frac{1}{2}\pi - \epsilon$ , (21.1) may be written

$$\sin^2 \frac{1}{2}\alpha = \mu \sin \eta \cos (\phi - \eta) \sin \frac{1}{2}\alpha / \sin (\epsilon + \frac{1}{2}\alpha),$$

and thus, with  $\sin \phi_0 = \mu^{-1}$ ,  $\sin^2 \frac{1}{2}\alpha \rightarrow \mu \sin \eta \cos (\phi_0 - \eta)$  as  $\epsilon \rightarrow 0$ .

Let APK, normal to V, meet V at P, and let PK = AP. Then BK = BA, and the ray BC is directed from K. Let AN, normal to U, meet V in N, and let AN =  $t$ . Then PAN =  $\eta$ , CKA = BAK =  $\phi - \eta$ , and ACK =  $\frac{1}{2}\pi - (\phi - 2\eta)$ . We have AK =  $2t \cos \eta = 2OA \sin \eta$ , and

$$\begin{aligned} AC &= 2OA \sin \eta \sin (\phi - \eta) / \cos (\phi - 2\eta), & CK &= AC \cos \eta / \sin (\phi - \eta), \\ AF &= AC \cos (\theta - \alpha) / \sin \alpha, & CF &= AC \cos \theta / \sin \alpha. \end{aligned}$$

Hence

$$AF - CF = AC \sin (\theta - \frac{1}{2}\alpha) / \cos \frac{1}{2}\alpha.$$

Let the optical path from Q to F via ABCF exceed that via AF by  $\Delta$ . Then

$$\Delta = \mu(AB + BC) + CF - AF = \mu CK - (AF - CF).$$

Thus

$$\Delta / AC = \mu \cos \eta / \sin (\phi - \eta) - \sin (\theta - \frac{1}{2}\alpha) / \cos \frac{1}{2}\alpha. \quad (21.3)$$

For small  $\eta$ ,  $t$  becomes  $OA(\eta + a\eta^3 + \dots)$ , AC becomes

$$2OA(\eta \tan \phi + b\eta^2 + \dots),$$

and the terms on the right of (21.3) become

$$\mu / \sin \phi + c\eta + \dots, \quad \sin \theta + d\alpha + \dots,$$

where  $a, b, c, d$  are independent of  $\eta$ . To first order,  $\alpha$  is given by (21.2). With  $\eta < 10^{-3}$  radian, we may neglect the terms containing  $a, b, c, d$ , and then, with  $\sin \theta = \mu \sin \phi$ ,

$$\begin{aligned} \Delta &= 2OA \tan \phi \cdot \eta(\mu / \sin \phi - \sin \theta) = 2OA \cdot \eta(\mu - \sin \theta \sin \phi) / \cos \phi \\ &= 2\mu t \cos \phi = 2OA \cdot \eta(\mu^2 - \sin^2 \theta)^{\frac{1}{2}}. \end{aligned} \quad (21.4)$$

We have

$$AF = \frac{2OA \sin \eta \sin (\phi - \eta) \cos (\theta - \alpha)}{\cos (\phi - 2\eta) \sin \alpha}.$$

To the first order,  $AF = 2OA \tan \phi \cos \theta \cdot \eta / \alpha$ , and, by (21.2), unless  $\theta$  be nearly  $\frac{1}{2}\pi$ ,

$$AF = \frac{OA \sin \phi \cos^2 \theta}{\mu \cos^2 \phi} = OA \sin \theta \frac{1 - \sin^2 \theta}{\mu^2 - \sin^2 \theta}. \quad (21.5)$$

When we put  $\theta = \frac{1}{2}\pi$  in the accurate value of AF, we find  $AF = AC$ .

We note that AF is positive or negative as  $\theta$  is positive or negative, or as OAQ is obtuse or acute.

For a given  $\theta$ ,  $AF/OA$  is constant, and  $F$  lies on a fixed straight line through  $O$ . Let  $OH$  be perpendicular to  $AF$ , and let  $FOH=\omega$ . Now  $AH=OA \sin \theta$ ,  $OH=OA \cos \theta$ , and thus, by (21.5),

$$\tan \omega = \frac{AH-AF}{OH} = \tan \theta \left( 1 - \frac{1-\sin^2 \theta}{\mu^2 - \sin^2 \theta} \right) = \tan \theta \frac{\mu^2 - 1}{\mu^2 - \sin^2 \theta}. \quad (21.6)$$

## § 22. Results for $\mu=1$ .

If  $\mu=1$ , then  $\phi=\theta$  and  $\alpha=2\eta$ , and the approximate results become

$$\Delta = 2t \cos \theta, \quad AF = OA \sin \theta, \quad \omega = 0.$$

By § 10, the exact values are

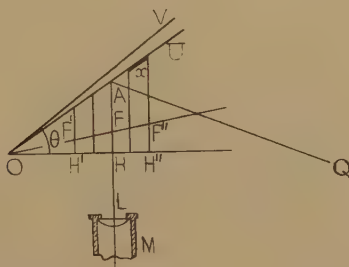
$$\Delta = 2t \cos \theta, \quad AF = OA \sin \theta - t \cos \theta, \quad \omega = \eta.$$

The small differences have no practical significance.

## § 23. Method.

The flanking plates do not change the spacing of the fringes or the direction of the straight line  $OF$  (fig. 13) on which they lie. In this section, we may neglect them.

Fig. 14.



The faces  $U$ ,  $V$ , of the wedge are vertical. Light from a source  $Q$  (fig. 14) is reflected at the wedge, and enters the microscope  $M$ . The axis  $AL$  of  $M$  makes  $\theta$  with the normal to  $U$ . When  $M$  is adjusted, vertical fringes are seen. The path difference changes by  $\lambda$ , the wavelength in air, from one fringe to the next. Let  $x$  be the distance along  $U$  which corresponds to a change of  $\lambda$  in  $\Delta$ . Then, by (21.4),

$$\lambda = 2\eta x(\mu^2 - \sin^2 \theta)^{\frac{1}{2}}. \quad (23.1)$$

The projection of  $x$  on  $OH$ , which is perpendicular to  $AL$ , is  $x \cos \theta$ . If the fringe spacing, as seen on  $S$ , the eye-piece scale of  $M$ , be  $\xi$ , then  $\xi = kx \cos \theta$ , where  $k$  is constant.

When the wedge is of air, the fringes, for a given  $\theta$ , lie on  $OH$  (see § 22); when it is of liquid they lie on  $OF$ , and  $FOH=\omega$ . With water of index  $4/3$ , and for  $\theta=30^\circ$ ,  $45^\circ$ ,  $60^\circ$ ,  $\omega=16.4^\circ$ ,  $31.3^\circ$ ,  $52.7^\circ$ . Hence the fringes along  $OF$  cannot be in perfect focus over the whole of  $S$ . If  $F'F''$ , the greatest width on  $OF$  which can be imaged on  $S$ , be small compared with the distance from  $F$  to  $L$ , the objective of  $M$ , the fringes  $F'$ ,  $F''$  will be

only a little out of focus when F is in focus. If  $F'F = FF''$ , the distance on S covered by  $F'F$  will be a little greater than that covered by  $FF''$ , because  $F'L$  is less than  $F''L$ . But the distance on S covered by  $F'F''$  will be very nearly equal to that covered by  $H'H''$ , when H is in focus. This result is used in § 25.

If  $x_A, x_W$  be the values of  $x$  for air and for water,

$$\xi_A = kx_A \cos \theta, \quad \xi_W = kx_W \cos \theta.$$

Then, by (23.1)

$$\lambda k \cos \theta = 2\eta \xi_A (1 - \sin^2 \theta)^{\frac{1}{2}} = 2\eta \xi_W (\mu^2 - \sin^2 \theta)^{\frac{1}{2}}. \quad (23.2)$$

Hence

$$(\mu^2 - \sin^2 \theta) / (1 - \sin^2 \theta) = (\xi_A / \xi_W)^2 \equiv q^2, \quad (23.3)$$

and

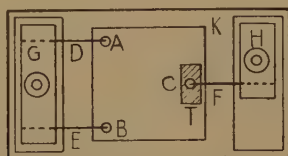
$$\mu = \{q^2 - (q^2 - 1) \sin^2 \theta\}^{\frac{1}{2}}. \quad (23.4)$$

By (23.2),  $\xi_A = k\lambda/2\eta$ , and so  $\xi_A$  is independent of  $\theta$ , as already shown in § 12.

#### § 24. *Experimental Details.*

One flanking plate rests on three knobs A, B, C (fig. 15) projecting from the vertical backboard K. Black varnish on the back of this plate reduces reflexion there. One end of the front plate rests on the

Fig. 15.



back plate. At the other end, the plates are separated by T, a small piece of tin-foil. The plates are pressed together by springs D, E, F. Springs D, E are fixed to a metal slab G, and F to a slab H. The slabs are secured to the backboard by screws. The springs touch the front plate at points opposite to A, B and C. If the opposed surfaces of the plates be very nearly plane, the fringes will be nearly straight and nearly equally spaced. The system of three supports and three springs prevents strains, which may easily make great changes in the fringes. A little care in getting contact (dust motes cause trouble) between the plates at the AB-end, and in adjusting the points of contact of the springs, may be needed if the fringes are to be vertical. A fringe spacing of 0.5 to 1 mm. is convenient.

The plates must be clean and free from grease, so that, when the time comes for it, water will spread evenly between them from a drop placed on the top edge. A little acetone is put into a dish, and the appropriate face of each plate is dipped into it for a few seconds. The acetone evaporates rapidly. The plates are then assembled.

The backboard is fixed to a circular table turning in a bearing having a vertical axis, which passes through the centre of the wedge. The bearing



is fixed to an optical bench. On the edge of the table are a zero and marks for  $\pm 30^\circ$ ,  $\pm 45^\circ$ ,  $\pm 60^\circ$ . The microscope is mounted on the bench; as nearly as may be, its axis is parallel to that of the bench and passes through the centre of the wedge.

The plates are set normal to the bench axis by the method of § 18. A pointer, adjustable relative to the base carrying the table bearing, is then set to the zero of the table, and is clamped to the base.

We suppose that the thick end of the wedge is throughout towards the right of the observer. For a given  $|\theta|$ , there are two positions of the table. In Position 1, corresponding to figs. 13, 14, the thin end is nearer the microscope than the thick. In Position 2 the reverse is the case. In Position 1 the source is to the right, and in Position 2 to the left, of the observer. For Position 1, the fringes are nearer to, and for Position 2 are farther from, the observer than is the wedge.

Fig. 16.

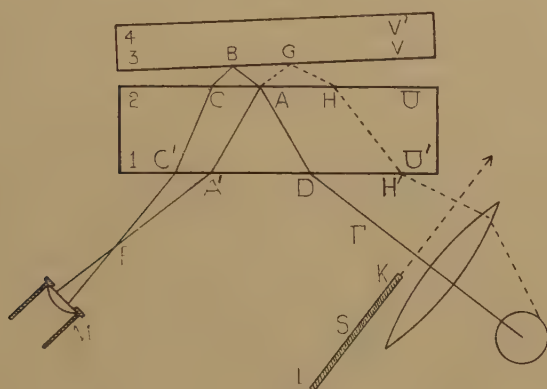


Fig. 17.



Fig. 18.



With the table set for, say  $+30^\circ$ , the source is put about 50 cm. from the wedge and so that the observer, looking along the axis of the microscope, sees the source by reflexion. If a converging lens of focal length  $f$ , (10—15 cm.) and 5 cm. in diameter, be placed between source and wedge at distance  $f$  from the source, fringes will cover a large area and will fill the field of view of the microscope.

There are four reflectors, viz.,  $U'$ ,  $U$ ,  $V$ ,  $V'$ , where  $U'$ ,  $V'$  are the outer faces of the flanking plates; they are numbered 1, 2, 3, 4 in fig. 16. The microscope  $M$  can receive light from the (extended) source by reflexion at each of the four faces; multiple reflexions are, for the moment, disregarded. Let  $I_1$ ,  $I_2$ ,  $I_3$ ,  $I_4$  be the respective intensities at the micrometer scale. The fringes depend upon  $I_2$ ,  $I_3$ . When  $V'$  has its natural surface, and the wedge is of air,  $I_2$ ,  $I_3$  are comparable with  $I_1$ ,  $I_4$ , and the fringes are not swamped by the light reflected at  $U'$  and  $V'$ . They can, in fact, be received on a screen of ground glass when the wedge is in Position 1. With the water wedge,  $I_1/I_2$ ,  $I_4/I_2$ ,  $I_1/I_3$ ,  $I_4/I_3$  may be

five or more, and the fringes are lost in the general glare. Black varnish on  $V'$  greatly reduces  $I_4$  and makes the fringes visible. But they are still weak, for the strong  $I_1$  remains.

The beam due to reflexion at  $U'$  can be prevented from entering  $M$  by a screen  $S$ , with vertical edges  $K, L$  (fig. 16). Let  $\Gamma$  be the single ray in the principal plane which gives rise to the rays  $A'F, C'F$  which "interfere" at  $F$ . Let  $S$  be moved in the direction of the arrow across the wide incident beam, starting from a position in which it intercepts no light. Until  $S$  begin to cut off light which could, after reflexion at a narrow patch round  $A'C'$ , enter  $M$ , the fringes are weak; they are small sinusoidal variations of brightness. But, as soon as the edge  $K$  cuts off all such light, the fringe  $F$  corresponding to  $\Gamma$  stands out strongly as a narrow dark band and so continues until  $K$  reaches  $\Gamma$ , when, of course, the fringe disappears in the shadow of  $S$ . Fig. 17 gives the view in the (inverting) microscope. In the bright region to the left of  $K'$ , the shadow of  $K$ , are three or four strong bands, dark and narrow; the fringes further to the left are those just described as "weak". The number of strong dark fringes depends upon the distance  $DA'$ . As  $K'$  advances in the direction of its arrow, the positions of the fringes do not change, but the first three or four to the left of  $K'$  are always strong.

If  $S$  be moved further, so that  $K$  and  $L$  are on opposite sides of  $\Gamma$ , there will be no fringe at  $F$ , nor will a fringe appear until  $L$  has passed  $\Gamma$ . But, before the screen is clear of  $\Gamma$ , light will have reached  $A'C'$ ; fig. 18 gives the view in the microscope. To the right of  $L'$  there is a bright space, three or four fringes wide, free from fringes. Further to the right there are "weak" fringes like those at a distance from  $K'$  in fig. 17.

The bands are narrow, because, with the oblique incidence employed the rays suffering 2, 3, . . . reflexions at  $V$  (fig. 16) are strong enough to influence the distribution of intensity in the fringe pattern. With semi-silvered surfaces, very narrow bands may be obtained. The bands obtained by the use of  $S$  do not reach that standard, but they are markedly narrow compared with their spacing.

The narrow bands are dark, and not bright, because there is a relative reversal of phase for the reflexions at  $A$  and  $B$ .

Let a ray  $FA'A$  be refracted along  $AG$  and, after reflexion at  $G$ , give rise to  $GHH'$ . Then a ray  $H'HGA$  gives rise to the two rays  $A'F, C'F$ . The corresponding fringes are narrow and are brighter than the background (see § 7). They may be seen with the air wedge, if the screen  $S$  be suitably adjusted. The source  $Q$  must be powerful.

### § 25. *Practical Example.*

The observations were made by C. G. B. Garrett and G. F. C. Searle. A sodium vapour lamp was used. The readings of the dark fringes on the eye-piece micrometer scale (centre at 50) were combined as below for the case when the wedge, of air, was in Position 1, *i. e.* with its thick end

towards the source, and  $\theta$  was  $30^\circ$ . Here  $m$  is the number of micrometer spaces covered by  $n$  fringe spaces. Hence  $m/n = \xi$  (§ 23).

Fringe.	Microm.	Fringe.	Microm.	$m$ .	$n$ .	$m/n$ .
1	1.3	18	96.7	95.4	17	5.612
2	6.7	17	91.0	84.3	15	5.620
3	12.3	16	85.3	73.0	13	5.615
4	18.0	15	79.4	61.4	11	5.582
5	23.5	14	73.9	50.4	9	5.600
6	29.2	13	68.2	39.0	7	5.571
7	34.8	12	62.5	Mean $\xi_A = 5.600$		
8	40.4	11	56.9			
9	45.8	10	51.4			

The values of  $m$  become smaller, and those of  $m/n$  less reliable, as  $n$  diminishes; results for  $n < 7$  are discarded. The results for Positions 1 and 2 were as follows:—

	Position 1.			Position 2.		
$\theta$ .....	$30^\circ$	$45^\circ$	$60^\circ$	$30^\circ$	$45^\circ$	$60^\circ$
$\xi_A$ .....	5.600	5.567	5.597	5.629	5.561	5.674

The general mean is  $\xi_A = 5.605$ . By the theory,  $\xi_A$  is independent of  $\theta$ . The differences are perhaps due to errors of focusing.

The observations with air being ended, water was put between the plates, and readings were taken over, roughly, the range 25 to 75 of the micrometer scale. The readings, treated just as those for air, gave the results:—

$\theta$ .....	$30^\circ$	$45^\circ$	$60^\circ$	
$\xi_W$ .....	3.891	3.482	2.738	Position 1
$\xi_W$ .....	3.882	3.467	2.792	Position 2
Means .....	3.886	3.474	2.765	

By (23.3),  $q = \xi_A / \xi_W$ . With  $\xi_A = 5.605$ , we have, for  $\theta = 30^\circ, 45^\circ, 60^\circ$ ,  $q = 1.4424, 1.6134, 2.0271$ . Then, by (23.4),  $\mu = 1.345, 1.342, 1.333$ , with mean 1.340.

XXXVIII. *New Contributions to Interferometry.* Part VII.—*Monochromatic and White-light Multiple Beam Interferometric Studies on Mica.*

By S. TOLANSKY, Ph.D., D.I.C., Manchester University\*.

[Received May 14, 1945.]

[Plates III.—V.]

*Abstract.*

New multiple beam interferometric studies are made on thin sheets of mica (some 1/100th millimetre thick) coated on both sides with high reflecting silverings. Three multiple beam methods are employed: (a) monochromatic Fizeau fringes, (b) new white-light fringes named "fringes of equal chromatic order," (c) new monochromatic non-localized fringes. The Fizeau fringes reveal marked cleavages. It is proved that both mica surfaces are highly distorted but that the contours follow each other exactly, so that the mica thickness remains constant (allowance being made for local cleavage steps). The surfaces are thus effectively parallel. Local variations in chemical composition reveal themselves. Much more information is given by the fringes of equal chromatic order. From these, the total mica thickness, height and absolute direction of cleavage steps, formation of surface topography, birefringence, etc., can be determined. The sensitivity is 60 per cent. better than in the case of the air film fringes of equal chromatic order described in Part V. Changes in thickness of the mica sheet are found to be discontinuous, the observed steps varying from one "molecule" upwards. A single "molecular" lattice (20 A.U.) can now be evaluated with certainty. The birefringence of the mica used has the value 0.0046. The birefringence doubling of the fringes of equal chromatic order is shown both theoretically and experimentally to be independent of the thickness of the mica. This is an unexpected result. A birefringence as low as 0.0001 can be measured by the new method. A description is given of the method for using the non-localized multiple beam monochromatic fringes formerly obtained with a Fabry-Perot interferometer. This leads to fringes of abnormally high dispersion which reveal (a) cleavage steps with high precision, (b) birefringence effects, (c) local variations in mica quality. The combination of the three independent interference methods is of considerable power.

---

*Introduction.*

IN Part V. a description is given of the application of what are here called "fringes of equal chromatic order" to the study of the topographical features of a single mica surface. In this section an account

---

\* Communicated by the Author.



is given of interferometric studies on mica by three distinctly different multiple beam methods, applied to very thin slips of mica (of the order of 1/100th millimetre thick). The methods involve (a) Fizeau fringes with monochromatic light, (b) the new fringes of equal chromatic order obtained with white light, (c) the new multiple beam non-localized ring fringes described by the writer earlier. These three separate methods, each very sensitive, together form a powerful analytical combination and compliment each other in an interesting manner. For method (a) gives information about extended areas, (b) refers to a line section, and (c) gives restricted information about either small local areas (crude points) or extended areas.

It has been shown elsewhere that the topographical features on mica are usually highly complex, and some hesitation was felt before attempting to extend the studies to mica slips silvered on *both* sides because of the anticipated increase in complexity due to the superposition of effects from both the mica surfaces. Surprisingly enough, in some respects the phenomena which appear with doubly silvered micas are in some cases of considerably less complexity.

The experiments reported here were made on thin rectangular slips of muscovite micas each some  $3 \times 4$  centimetres, and some 1/100th millimetre thick. These thin slips were sensitive to air shocks, and it was found convenient to mount them between glass plates after silvering both sides by evaporation *in vacuo*. The silver deposited was of the usual high reflecting coefficient employed in all these multiple beam investigations. The clamping of the silvered mica between glass leads to the formation of secondary fringes, which can be made use of in a manner to be indicated later. In some investigations the mica was clamped between metal plates suitably drilled to give an opening, and in this case the secondary fringes do not appear. Most of the work was carried out with glass-enclosed specimens, for this considerably retards the destruction of the reflecting properties of the silver film, which always sets in rapidly in Manchester with films exposed to the air.

#### *Observations with Fizeau Fringes.*

The usual arrangement for producing multiple beam Fizeau fringes was adopted. The doubly silvered mica was illuminated with parallel monochromatic light (green mercury) at normal incidence, using a small source which was at the principal focus of a lens. Since mica is doubly refracting the incident light was first suitably plane polarized to reduce the complexity. The interference patterns formed are of much interest, and are even more striking to the eye if the mercury arc source is not filtered for monochromatism.

The monochromatic patterns for some samples are shown in Pl. III., also in Pl. IV. (a) and (c). The photographs III. (a), III. (b), IV. (a), IV. (c), on the same scale, represent variously some 3 to 4 square centimetres of mica surface. Pl. III. (c) and (d), are six times enlargements of portions of Pls. III. (b) and Pl. IV. (c) respectively.

It can be seen that there exist large areas of uniform tint separated by distinctive cleavage lines. In Pl. IV. (a) these lines have diverged from the point of insertion of the needle used for starting the cleavage. The cleavage lines can be clearly seen on the mica surface with the aid of a lens. In Pl. III. (b) the roughly parallel lines are on one mica face, and the long line crossing these roughly perpendicularly is a cleavage on the other face. The sudden bending in all the lines simultaneously on Pl. III. (a) gives evidence of some extended discontinuity.

Within the uniform tint areas are features of interest. There are (a) secondary fringes of low visibility, (b) gradual changes in the depth of the tint, (c) marked definite striations and irregular discontinuous patches.

The origin of the secondary fringes is clear. They arise from interference between the silvered mica surfaces and the closely contacting glass supporting faces. They are therefore low visibility complexes arising from the crossed fringes from two silvered mica faces and two highly irregular unsilvered glass faces. They are simplified (this has been done in some cases) by introducing a piece of paper between the edge of one mica face and the glass, producing a relatively large angle wedge on this side, whereby effectively the secondary fringes which are visible refer to one mica face and one glass face.

The observations were extended by resting a *doubly* silvered piece of mica upon a *silvered* optical flat leading to sharp compound fringes. The mica acts as a filter leading to uniform tint areas. But within each area of transmission sharp fringes (due to multiple beams) are formed between the silvered mica surface and the silvered glass in contact with it.

Pl. III. (c) and (d), were taken with the aid of a combination of microscope and camera, the particular enlargement in these cases being 22 times.

The interpretation of these fringes is of interest. Uniform tint, when observed either with a filtered or unfiltered mercury arc (sensitivity is much increasing with mixed radiations), means that both faces of the mica must be very closely parallel. This must be true to within a small fraction of a wave. However, the secondary complexes prove that the surface of the mica is highly distorted. It follows that the distortions on the one surface are faithfully followed on the other. That such "parallel" distortions will still lead to uniform tint is evident from consideration of the formula  $n\lambda = 2\mu t \cos \theta$ . For  $t$  is postulated as constant, and the small variations in  $\theta$  due to the contours are entirely negligible. Thus it is clear why Fabry-Perot fringes (or the original Haidinger fringes) can be perfectly formed with very imperfect mica surfaces.

Considering now the two other characteristic features, namely gradual changes in depth of tint and striations, the following conclusions can be drawn. The gradual regular changes in tint occur without any apparent correlation with surface topography. It is reasonable to conclude that they represent variations in optical thickness arising from variation in chemical composition. It is known that micas are not quite homogeneous, and a slow regular substitution of one kind of atom by another would lead to a small change in refractive index, with a consequent slow

regular alteration in tint. It can be seen that the changes in tint are gradual in some cases and more abrupt in others.

Distinctly different are the striations. Some of these are shown in the enlarged section of Pl. III. (a) shown in Pl. III. (c), which represents an area on the mica of about six square millimetres. The striations are quite independent of the secondary surface fringes and are clearly inclusions of substance of refractive index differing from that of the surrounding medium. These are definitely not visible on the mica when examined (unsilvered) by simple transmission. They are probably not foreign occlusions in the usual sense, but rather to be considered as areas of mica with modified structure created during the laying down process. They tend to show a general orientation, being more or less parallel streaks. It is of interest to note that on Pl. III. (c) there are two sets of streaks crossing in different directions and very probably laid down at different times, *i. e.* occurring at different depths. No more precise information can be derived about these. They are present only in some specimens.

#### *Observations with Fringes of Equal Chromatic Order.*

Owing to the local parallelism of the sides of the mica, the Fizeau fringes are unable to give any information about the heights of the steps or, in effect, about the variations of total thickness of the mica produced by cleavage steps. It is just in this connection that the fringes of equal chromatic order (see Part V.) reveal their usefulness, and their application to mica will now be considered in detail.

The fringes of equal chromatic order for a line section, about 1.5 centimetres long, crossing over two of the cleavage lines of Pl. IV. (a), are shown on Pl. IV. (b). The fringes are double because of the birefringence of the mica, each component being plane polarized, mutually perpendicularly. The fringes are straight lines parallel to the spectrograph slit (curvature of the slit image being taken into account by comparison with a line spectrum). This is an additional proof that the faces of the mica are parallel between a pair of cleavage lines, for these fringes are clearly the equivalent of the Edser-Butler fringes given by a plane parallel film.

The fringes reveal many interesting properties, some of which are as follows. They reveal the direction of the cleavage step, that is, they show that the upper portion above the cleavage line is thicker than the lower. Taking  $\mu$  to be 1.584 at 5900 A.U., the metrical thickness (as distinct from the optical thickness) of one of the steps is 120 A.U., *i. e.* six "molecules."

Next, simple calculation enables the total thickness of the mica slip to be obtained. It is 1/59th millimetre.

It will be noticed that the fringes reveal a secondary variation in intensity. These are the secondary fringes of equal chromatic order formed between the silvered mica face and the optical flat (unsilvered) upon which it rests. The combination of the straight fringes with their superposed complex secondaries show, in a further striking fashion, that although the surface is highly distorted, yet the other mica surface follows



the same distortion faithfully. The evidence given by the Fizeau fringes of the existence of large uniform areas is clearly confirmed by the extent of the uniform fringes in Pl. IV. (b).

Consider now the fringes shown in Pl. III. (e), which were obtained with the mica sample shown in Pl. IV. (c). They represent a section about 1.4 centimetres in length, running vertically across the black rectangular patch in the middle left of Pl. IV. (c), *i. e.* the area of which Pl. III. (d) is an enlargement. The fringes on Pl. III. (e) are magnified about 1.4 times greater than those on Pl. IV. (c). The bright narrow streak across the middle of the dark patch corresponds to the small stepped fringe groups in the centre of the black pattern of Pl. IV. (c). The correspondence between the two systems of fringes can be exactly traced. Pl. IV. (d) refers to another section of Pl. IV. (c), this time taken in a more drastically torn region. The scale of enlargement is the same as in the previous pattern, whilst Pl. IV. (e) is an enlarged section of Pl. IV. (d). Pl. IV. (f), on the same scale as Pl. IV. (e), refers to still another sample, and is of equal interest.

A study of these complex patterns reveals many points. The thickness of the mica (sample I. (c)) is some 1/140th millimetre. It is quite clear that the surface is violently torn into small strips, but between cleavage lines the areas are still approximately parallel. The steps are randomly distributed both as to sign and amount in some areas, but in others follow a regular course. It is impossible even to attempt to allocate the steps to the two surfaces of the mica, the whole being confused. The fringe patterns, being strictly pictures of the optical thickness of the mica slip, are magnified in displacement effects by the factor  $\mu$  when compared with air film fringes of equal chromatic order. The sensitivity of the method, using doubly silvered mica, is thus 60 per cent. better than that using an air film.

Disregarding special local features to be considered later, it appears, from the examination of many specimens, that changes in the mica thickness are discontinuous. Lines which appear to slope gradually, appear under higher magnification to consist of discrete steps (see Pl. IV. (f)). In general, therefore (apart from the localized exceptions to be discussed later), the mica *does not exhibit wedges of continuously varying thickness*.

A frequent feature which can be seen in many places occurs wherein, with three adjacent strips, the two outer coincide in level, whilst the central fringe is displaced either to red or violet. This simply indicates the tearing out of a strip from a uniform area, or alternatively the production of a strip ridge on a plane surface, both outer sides being at the same level. The regular ascending series of steps, as in Pl. IV. (f) (and on many other examples), cannot be fortuitous. It is possible that those occasional single odd steps breaking up such sequences are attributable to the other side of the mica, and do not belong to the sequence. The tendency to cleave in such sequences is clearly related to the order in which the layers have been put down during the building up of the crystal.



The whole clearly marked step sequence in the centre of Pl. IV. (f) occupies only 800 A.U., *i. e.* corresponds to only 40 "molecules." Individual steps are of the order 11, 10, 4, 7, 9 "molecules" respectively, taking the molecular lattice as 20 A.U.

From the viewpoint of the sensitivity of the method, the measurements made on the fringes are of interest. The following table gives a series of measured steps in A.U., and below this the corresponding equivalent number of "molecules." The error does not exceed 4 A.U. and is probably less.

Step A.U.....	41	-41	79	-40	134	-137	-18	17	-133	127
"Molecules" ....	2	2	4	2	7(6½)	7	1	1	7(6½)	6(6½)

This table is of considerable interest. It is quite an optical achievement to measure down to one molecular lattice with such certainty. The known lattice spacing being 20 A.U., the closeness of the smaller steps to exact multiples of 20 affords satisfying confirmation of the precision of the method, agreeing with the estimated error derived from the consistency with which readings can be made. It is to be noted that in rare cases where a microphotometer could be employed this might be improved upon. It is further to be noted that the internal consistency confirms the earlier conclusion that the silver film contours the surface with very great exactness.

There are reasons to suppose that *half* integral steps might occur in cleavage of mica. The above evidence gives a certain amount of indecisive support to this view. Other measurements made lead to no more certain conclusion. This question must therefore be left open.

### *Birefringence of Cleaved Mica Sheets.*

The fringes of equal chromatic order permit of a determination of the birefringence. The refractive indices of various micas are known to differ according to composition, and values found with precision for any one sample need not necessarily be representative. The three refractive indices (*averages*) given by Rutley ('Mineralogy,' 23rd edition) are 1.560, 1.593, 1.600. The expected birefringence in the cleavage flakes is thus of the order of 0.007, a low value.

The doubling of the fringes is due to birefringence, and the relation between alteration in refractive index and wave-length is easy to derive. We have  $n\lambda = 2\mu t$ . For a given order ( $n$  constant) and a given thickness, differentiating gives  $nd\lambda = 2d\mu t$ . Hence the birefringence  $d\mu$  is given by  $d\mu = \mu d\lambda / \lambda$ . This can be measured over the spectral range by measuring the doublet separation  $d\lambda$ . The value of  $\mu$  can be estimated from the wave number separation between any two orders.

Measurement shows that  $d\mu$  is sensibly constant over the visible spectrum, and with the particular sample used has the value  $0.0046 \pm 0.0001$ .

Since  $d\lambda = d\mu \lambda / \mu$ , we arrive at a most unexpected result. The birefringence doubling, in A.U., is independent of the thickness of the mica

film. That this is indeed the case is clear from casual inspection of the various plates reproduced here, for these have been taken with varying values of film thickness.

This prediction of theory was put to an exact test by making measurements on two pieces of mica of different thickness cleaved from the same sheet. The mica thickness can be obtained with precision from the  $\Delta\nu$  separation of the orders. The two thicknesses were respectively,  $1.57 \times 10^{-2}$  and  $0.72 \times 10^{-2}$  millimetre, *i. e.* in the ratio 2.18:1. As closely as could be measured the  $d\lambda$  values at a given wave-length were identical. The value of the birefringence separately calculated from the two sets of observations was identical, namely 0.0046.

It is estimated that a birefringence of 0.0004 can be measured easily, whilst a value as low as 0.0001 could be measured by first polarizing the incident light and then separately measuring the positions of the fringes on the wave-length scale.

#### *Observations with Non-Localized Fringes of Equal Inclination.*

A third interferometric method has been developed for the examination of doubly silvered mica. This makes use of the non-localized multiple beam monochromatic circular fringes described earlier in Part I. (Tolansky, 1943):

It has long been recognized that suitably selected pieces of mica, when silvered on both sides, can produce *Fabry-Perot* rings if a small area is illuminated. The original discovery by Haidinger of the first known fringes of equal inclination (made in 1854) was made with mica. Rayleigh (1906) first deposited silver by cathodic sputtering on to the two faces of a slip of mica, and as a result was able to resolve two ring systems, due to the birefringence. This was developed further by Chinmayanandam (1919), and also by Schafer and Herbert (1926). More recently Pfund (1942) has extended these observations without apparently being aware of the earlier literature.

The non-localized fringes, formerly described by the author in Part I., are very closely allied to true Fabry-Perot rings, although the latter are at infinity. It was shown that with a small source (approximately a point source) a Fabry-Perot interferometer produces highly sharpened rings which extend out indefinitely into space, growing larger as the distance between the screen and source is increased. No lenses are required for the production of these fringes, and providing the separation of the interference surfaces is a small fraction of the distance between screen and source, there appear on the screen highly sharpened rings, with an intensity distribution very closely similar to that of true Fabry-Perot rings.

A characteristic of these rings is the very large diameters that can be obtained with even 1 millimetre separation between the surfaces. Rings many metres in diameter can readily be obtained with mica. Dispersions are available which are undoubtedly greater than those usually obtained by any other interference procedure. In fact, with thin slips of mica, so great

are the ring diameters that it is necessary to tilt off-centre to accommodate more than one order on a quarter-plate placed only 10 centimetres away from the source.

The optical arrangement employed is simple. A small aperture is placed before a green-filtered mercury arc. With a one-inch microscope objective a diminished bright image of this circular aperture is focussed onto the mica. The size of the image used was somewhat less than one-third of a millimetre in diameter. The mica is tilted to bring down higher orders into the field, and a photographic plate is held some 20 centimetres away from the mica. Successive orders given when an image of the source is thrown into the uniform tint area of Pl. IV. (a) are shown in Pl. V. (a). The light was first plane polarized suitably to cut out one of the ring systems. The photograph on the Plate is a *direct contact print, not an enlargement*. As already stated it is easy to obtain rings many metres in diameter.

The rings are remarkably sharp, and the enormous dispersion is evident, for with this particular source the change in thickness corresponding to a single order change ( $dn=1$ ) is 1700 A.U., and from the sharpness it is evident that a change of 1/100th order could be detected. Although monochromatic light is used it is of interest to point out that the method is no more sensitive than the white light procedure of the fringes of equal chromatic order. A fringe *setting* can be made to 1/1000th of an order, *i. e.* to within 2 A.U.

The original plate shows some slight evidence of irregularity in the fringes, but the effect is too small to be measured. These fringes add confirmation to the earlier conclusion as to parallelism of the faces.

The mica was moved until the small source crossed over one of the strongly marked cleavage lines on Pl. IV. (a). On crossing this line a sudden discontinuity appears in the non-localized fringe pattern. This is shown in Pl. V. (b). We have now two effective Fabry-Perot interferometers, side by side, differing slightly in gap. The cleavage line divides the converging beam of light. The result is that on one side of the line only half a cone of converging light is available, and therefore only half the ring system is formed. On the other side of the cleavage line the other half of the ring system, from the second adjacent Fabry-Perot interferometer appears, with a displacement in order corresponding to the difference in gap. By measuring the displacement in order a precision value for the step in the mica can be obtained.

This picture alone is ambiguous, for there is no indication as to whether or not orders have overlapped, *i. e.* as to whether or not an unknown integer is to be added to the observed fractional order difference. The ambiguity can be resolved in two ways. One is laborious, and one simple. The laborious way is to use different wave-lengths and then apply Benoit's method of exact fractions for the evaluation of the two Fabry-Perot gaps. The simple way is to compare the fringes with the fringes of equal chromatic order, which at once tell whether an integer must be added. In this particular case no integer is to be added. The fringes of equal



chromatic order also indicate in which *direction* the step has gone, which cannot be obtained by inspection from the monochromatic fringes on Pl. V. The one advantage of the non-localized fringes over the fringes of equal chromatic order is the very great dispersion they offer with the simplest of means.

Pl. V. (c) was taken over the same region as Pl. V. (b) but with the polarizing device removed. The birefringence effect is strikingly obvious. One system of rings is *circular*, the other *elliptical*. This effect was already indicated, although far less clearly, in the earlier work of Rayleigh on the *true* Fabry-Perot rings, and was found by Pfund, also with *true* Fabry-Perot rings. None of the earlier workers who studied the Fabry-Perot rings succeeded in discovering the step effects observed here with the non-localized fringes.

The mica sample of Pl. IV. (a) was replaced by the sample of Pl. IV. (c), and the results obtained from a randomly selected area are shown in Pl. V. (d). It has already been shown that this mica has a violently torn up surface, and the effect of this is rendered in the fringes. In Pl. V. (a) we are clearly covering an area which contains a number of discrete patches all falling within the image and extending *horizontally* over a sufficient length to encompass a considerable range of angles for each strip. Even more complex patterns arise in some instances, and a bewildering flicker of fringes crosses the field if the mica is moved slowly across the light source.

Thus we have three distinctly different methods of examining the mica, Fizeau fringes, fringes of equal chromatic order, and fringes which are non-localized, and have the properties of Fabry-Perot fringes. The fringes of equal chromatic order are more powerful and simpler to apply in detail than either of the other methods, but all three act in a complementary manner. Together the combination is of considerable power. Clearly the procedures described need not be restricted to mica. For example, selenite is now being investigated. There are many other possibilities. Modern plastics can be produced in thin transparent sheets, and these should prove suitable for study.

Finally, it may be pointed out that the fringes of equal chromatic order appear to be a powerful method for studying weak birefringence. A systematic examination of the birefringence of various micas, including muscovite, lepidolite, biotite, and phlogopite, samples of which have been obtained for the purpose, is now being undertaken.

My thanks are due to Prof. H. G. Cannon, F.R.S., for the loan of an excellent micro-camera.

#### References.

- Chinmayanandam, Proc. Roy. Soc. A, xcv. p.176 (1919).  
Pfund, A., Journ. Opt. Soc. America, xxii. p.382 (1942).  
Rayleigh, Lord, Phil Mag. xii. p.489 (1906).  
Tolansky, S., Phil. Mag. vii. p.555 (1943).



XXXIX. *A Method for the Determination of Crystal Cuts by applying the Reflection of X-rays from a known Lattice Plane.*

By V. PETRŽILKA and J. BENEŠ \*.

[Received January 21, 1946.]

[Plate VI.]

THE purpose of this paper is to develop a method by means of which crystal cuts can be determined or controlled by the measurement of two angles, *i. e.* to fix the position of a cut surface by the measurement of two angles formed by the cut plane and the known lattice plane or the crystallographic axis of the crystal formed by the cut plate. The measurement of these two angles can be easily made with a precision of  $\pm 1'$ .

We come up against this problem in the manufacture of piezoelectric oscillators and resonators. The exact retention of a prescribed cut gives, in the case of piezoelectric oscillators and resonators, much better results in that it keeps, with maximum precision, the oscillations of a given frequency constant, with respect to the temperatures, *i. e.* they have a minimum temperature coefficient. Many papers in radio-communication have been published mentioning this requirement and dealing with the precise determination of prescribed crystal cuts.

If it is possible to use well formed crystals, the determination of crystal cuts can be made with a precision of  $\pm 6'$  by means of optical goniometers. This simplest method consists in a determination of the angle between a natural crystal surface and a crystal cut. A somewhat different method was used by A. V. Shubnikov<sup>(1)</sup>, who also relates the cut plane to a natural crystal surface, in the case of quartz to the plane  $R$  or  $r'$ , but he uses a 3-footed micrometer desk and libela.

It is not possible to use the above methods in the case of small or not-well-formed crystals, and these methods too have the disadvantage that, after the cutting out of the plate or a definite marking on the plate, they do not allow for a subsequent control of the cut. In such a case we can however make use of X-rays.

We find a note on the use of X-rays for this purpose in a paper published in 'Electronics'<sup>(2)</sup>. In this periodical the manufacture of quartz oscillators made by the General Electric Co. is described, and it is mentioned that for the determination of crystal cuts with a minimum temperature coefficient, X-rays were used. Details of this method are not mentioned any further. Burns<sup>(3)</sup>, too, quotes expressly that the determination of cuts for quartz rods with a minimum temperature coefficient was made in

---

\* Communicated by Nora Wooster, M.A., Ph.D.

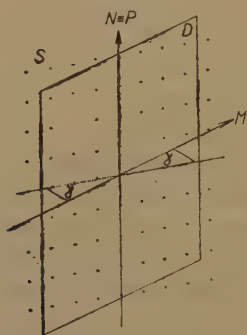
a spectrograph for X-rays with a precision of  $0.25^\circ$ . Further details are also omitted here. A more detailed description can be found in papers of Arkharov and Averkiev <sup>(4), (5)</sup>, who made use of the reflection of X-rays from a crystal fixed in a goniometric equipment. To complete the list we should also mention the papers of Majima and Togino <sup>(6)</sup>, who determine the crystal cuts from Laue-diagrams.

### 1. Principle of the Method used.

The principle of the method described in this paper consists in a determination of the deviation of the cut plane of the plate from the known lattice plane. The deviation is determined in such a way that in the cut plane of the plate two directions perpendicular to each other are chosen, and two angles are measured between these two directions and the lattice plane. For the measurement of the mentioned two angles the Seemann's spectrograph is used, which was for this purpose especially adapted.

The cut plane of the plate will be denoted by D. In this plane two characteristic directions perpendicular to each other, denoted M, N, are chosen. In the case of a rectangular plate these two directions can be, *e. g.*, parallel to its edges. According to the cut of the plate two special positions of the plane of the plate with respect to the lattice plane S can occur. Fig. 1 indicates a special case, in which the line of intersection

Fig. 1.



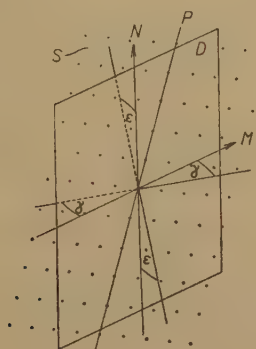
P of the cut plane D of the plate coincides with one of the two characteristic directions M and N. In fig. 2 the case is shown, in which the line of intersection P does not coincide with either of the characteristic directions M and N. Let us put through these characteristic directions two planes perpendicular to the lattice plane S. The lattice plane S and the cut plane D of the plate form in these planes two angles, which are in fig. 2 denoted by  $\gamma$  and  $\epsilon$ .

The plate is put in Seemann's spectrograph in such a way, that the characteristic direction N is parallel or coincides with the vertical direction of the knife-edge in Seemann's spectrograph, *i. e.* with the axis, round which the spectrograph swings. The direction N will be called vertical

and the deviation of the lattice plane S from this direction the vertical one. The direction M perpendicular to the direction N lies in the horizontal plane. Let us call this the horizontal direction, and the deviation  $\gamma$  the horizontal deviation.

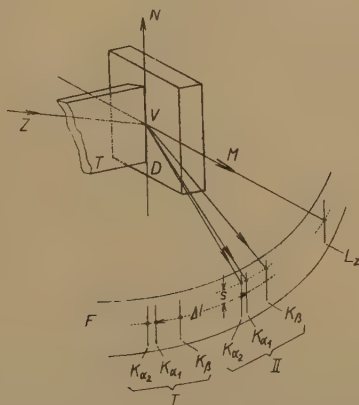
If we know exactly the position of the line of intersection P of the lattice plane with the cut plane of the plate, in other words if the plate has a horizontal deviation only (fig. 1), we can measure this deviation in the usual way by Seemann's spectrograph. The plate is placed in the

Fig. 2.



spectrograph so that the knife-edge is parallel with the characteristic vertical direction N, which coincides in this case with the line of intersection P. The determination of the angle  $\gamma$  is described in paragraph 2 of this paper.

Fig. 3.



If the position of the line of intersection P (fig. 2) is not known, it is necessary to adapt Seemann's spectrograph for the determination of the deviations  $\gamma$  and  $\epsilon$ . For this purpose the knife-edge T (fig. 3) is provided with a fine-point scratch V. During the exposure of the film F the scratch V shows itself as a small darker point on the spectral lines. The plate is put in the spectrograph in such a manner that the characteristic

direction  $N$ , denoted on the plate, is parallel or coincides with the sharp edge of the knife  $T$ . The measurement of the angles  $\gamma$  and  $\epsilon$  is described in paragraph 3 of this paper.

Finally, it is necessary to consider the case in which, by putting the plate in the spectrograph, the characteristic direction  $N$  shows a deviation  $\psi$  from the direction of the knife-edge  $T$ . This deviation can originate from an incorrect placing of the plate in the spectrograph. This deviation and the error of measurement will be discussed in paragraph 4 of this paper.

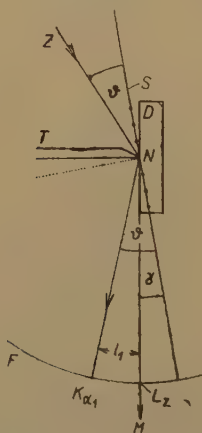
In paragraph 5 the method is described which makes it possible, when the angles  $\gamma$  and  $\epsilon$  are measured, to grind the plate so that it has the prescribed position of the line of intersection  $P$  and, in the plane perpendicular to the line of intersection  $P$ , the prescribed deviation  $\delta_0$  from the lattice plane  $S$ .

The application of the described method in the manufacture of piezo-electric oscillators is mentioned in paragraph 6.

## 2. The Plate has the Horizontal Deviation only.

If we look at Seemann's spectrograph from above, the horizontal deviation  $\gamma$  lies in the plane of the paper and the vertical direction  $N$  stands perpendicularly to the plane of the paper. Let us suppose that the characteristic direction  $N$  coincides with the line of intersection  $P$  of the lattice plane  $S$  and cut plane  $D$  of the plate in the crystal. We

Fig. 4.

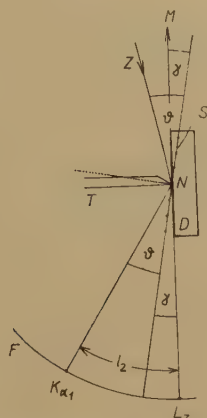


put the plate in the spectrograph so that the direction  $N$  is parallel with the knife-edge. There are two ways in which we can put the plate in the spectrograph. One way is that the direction  $N$  points up, the other way—if we rotate the plate through  $180^\circ$ —that the direction  $N$  points down. According to this fact, the position of the lattice plane  $S$  is reversible, as shown in figs. 4 and 5. On the film  $F$  the line  $L_2$  represents the track of the radiation  $Z$ , which goes during the swings of Seemann's spectrograph directly along the plate between the plate and sharp knife-edge.



We shall call it the main line. The line  $K\alpha_1$ , if, for example, we are using the K-series of the characteristic radiation of copper, has the  $l_1$ — or  $l_2$ — distance from the main line. If we denote  $\vartheta$ , the Bragg angle,

Fig. 5.



corresponding to the lattice plane S,  $\rho$  the radius of the spectrograph, *i. e.* the distance between the knife T and the film F, we can write, according to fig. 4.

$$\vartheta - \gamma = \frac{l_1}{2\pi\rho} 360^\circ, \quad . . . . . (1)$$

and, according to fig. 5,

$$\vartheta + \gamma = \frac{l_2}{2\pi\rho} 360^\circ. \quad . . . . . (2)$$

The precise determination of the distance  $l_1$  or  $l_2$  is difficult because of the fact that the line  $L_2$  is usually strongly exposed in comparison with the line  $K\alpha_1$ . The practical determination of the deviation  $\gamma$  can be done in another way. There are two possibilities:—

(a) We expose the same film, in both positions of the plate which are shown in figs. 4 and 5, to lines of the copper K-series; we measure the distance  $\Delta l_{12} = l_2 - l_1$  between them, and we determine the deviation  $\gamma$  from the formula

$$\gamma = \frac{l_2 - l_1}{4\pi\rho} 360^\circ = \frac{\Delta l_{1,2}}{4\pi\rho} 360^\circ \quad . . . . . (3)$$

according to the equations (1) and (2).

(b) We make the exposure either in the position of the plate given in fig. 4 or in the position of the plate given in fig. 5 and we expose on the same film the reflection from the plane which coincides with the lattice plane, *e. g.* the reflection from the natural face of the crystal, from which the plate was cut. According to fig. 6 we get

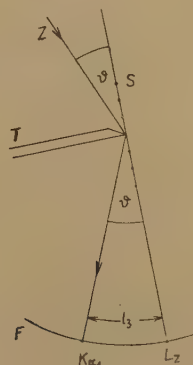
$$\vartheta = \frac{l_3}{2\pi\rho} 360^\circ. \quad . . . . . (4)$$

Substituting into equations (1) or (2) we obtain

$$\gamma = \frac{l_3 - l_1}{2\pi\rho} 360^\circ = \frac{l_2 - l_3}{2\pi\rho} 360^\circ. \quad (5)$$

In both cases (a) and (b) we use with advantage Seemann's spectrograph, adapted as shown in fig. 3. If the characteristic direction N coincides with the line of intersection P, as assumed, and if we place the plate correctly into the spectrograph, so that the direction N is parallel

Fig. 6.



with the knife-edge T, then the point on the main line  $L_z$  and the points on the system of the lines I. and II. (fig. 3) must lie in a straight line, the angle  $\epsilon$  being zero. The use of the sharp edge, with the scratch V, is in this case advantageous, giving the possibility of experimental control of the main physical assumptions involved in formula (3) or (5).

Figs. 4 and 5 show that it is possible to measure in this way the deviations  $\gamma$  as long as they are smaller than the Bragg angle  $\vartheta$  for the given lattice plane S. If these deviations are greater it is necessary to select another lattice plane lying closer to the cut plane.

### 3. The Plate has both Horizontal and Vertical Deviation.

If the line of intersection P of the plane D and of the lattice plane S does not coincide with one of the characteristic directions M and N, then the plate has not only the horizontal deviation  $\gamma$  but also the vertical deviation  $\epsilon$ , as is shown in fig. 2. The deviation  $\epsilon$  can be caused either by a wrong cut of the plate from the crystal or by a wrong placing of the plate in the spectrograph, so that N does not coincide with P. If we put such a plate in Seemann's spectrograph, adapted as shown in fig. 3, and if the characteristic direction N is parallel with the knife-edge T, the X-ray beam coming from the source Z and going through the scratch V is reflected as shown in fig. 7. The plane put through the characteristic direction M perpendicularly to the plane D of the plate is denoted H. The reflection of X-rays from the lattice plane S takes place in a plane  $\Sigma$  perpendicular to the plane S. Fig. 7 shows the beam  $Z_0$  going through the scratch V of the knife T, being reflected for the line  $K\alpha_1$  from the lattice plane S at the angle  $\vartheta$  and falling on the film F at the point B on the



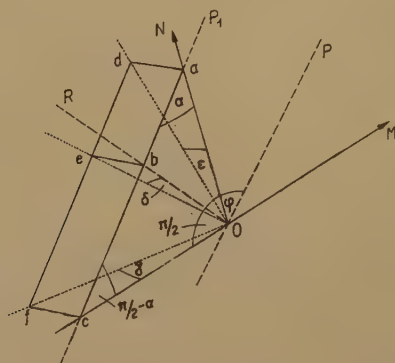
If we substitute  $\sin \vartheta$  according to this equation in equation (8) we get the formula (7). Calculating  $\sin \epsilon$  from equation (7) we get two values, from which only the one which has a positive sign before the root, namely,

$$\sin \epsilon = -\frac{1}{2A} + \frac{1}{2A} \sqrt{1 + 4A^2} = \frac{s}{2\rho \sin \vartheta} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

has any practical meaning, because we always select a lattice plane  $S$  such that the vertical deviation  $\epsilon$  may be small.

The formulas (6) and (7) or (10) give the possibility of determining both the horizontal and the vertical deviation. If we know both these deviations, we can determine the position of the line of intersection  $P$ . We denote by  $\phi$  the deviation of the line of intersection from the direction  $N$ , as shown in fig. 8, and we put through the straight line  $P_1$  parallel to

Fig. 8



$P$  a plane perpendicular to the lattice plane  $S$ . In addition to that plane we put another plane perpendicular to  $P$  and  $S$  through the straight line  $R$  perpendicular to  $P$ . In this way we get right-angled triangles  $ado$ ,  $beo$ ,  $cfo$ . According to these facts we can write

$$\begin{aligned} ad = eb = oa \sin \epsilon = ob \sin \delta = oa \sin \phi \sin \delta, \\ eb = cf = oc \sin \gamma = ob \sin \delta = oc \cos \phi \sin \delta. \end{aligned}$$

From these equations we get the equations for  $\phi$  and  $\delta$

$$\tan \phi = \frac{\sin \epsilon}{\sin \gamma} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$\sin^2 \delta = \sin^2 \gamma + \sin^2 \epsilon \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Figs. 9 and 10 (Pl. VI.) show the enlarged pictures of photographs taken by Seemann's adapted spectrograph. Fig. 9 (Pl. VI.) shows the K series of copper (denoted with II.) which has its origin in the reflection from the lattice plane of a quartz plate without a vertical deviation. Fig. 10 (Pl. VI.) shows also the K series of copper (denoted with II.) which has its origin from the lattice plane of a quartz plate with a vertical deviation. It is the reflection from the lattice plane parallel with the quartz face  $R$ . On both pictures by the series I. is denoted the K lines of copper reflected from the natural face  $R$  of quartz.





This formula shows that the wrong position of the plate in the spectrograph has no great influence on the determination of the horizontal deviation  $\gamma$ .

The wrong position has a greater influence on the vertical deviation  $\epsilon$ . According to fig. 11 and the formula (8) it follows that

$$s' = 2\rho \sin \vartheta' \epsilon_1.$$

Because

$$e = OV \sin \epsilon = OV_1 \sin \epsilon_1 = OV \frac{\sin \phi}{\sin (\phi + \psi)} \sin \epsilon_1.$$

it follows that

$$\sin \epsilon = \frac{\sin \phi}{\sin (\phi + \psi)} \sin \epsilon_1. \quad (16)$$

If, for example,  $\phi = \psi$ , then  $\sin \epsilon = \frac{1}{2} \sin \epsilon_1$ . In such an unfavourable case the determination of  $\epsilon$  is very inaccurate; because we are measuring for  $\epsilon$  almost double its value. This case occurs practically only when angles  $\phi = \psi$  are equal to  $1^\circ$  or  $2^\circ$ . Then is  $\epsilon$ , as shown by equation (11), so small that it does not matter if its determination is inexact, as will be shown by the numerical example. If *e.g.* the horizontal deviation is  $\gamma = 7^\circ$ , then the angle  $\phi = \psi = 1^\circ$  corresponds, according to the equation (11), to  $\epsilon = 0^\circ 7'$ . By measurement we get  $\epsilon_1 = 0^\circ 15'$ , as results from the equation (16). In practice we find that it is not necessary to commit, in adjusting the plate in the spectrograph, an error greater than  $\psi = 1^\circ$ . On the contrary it is possible after short training to obtain an average error smaller than  $\psi = \pm 1/3^\circ$ . We have only to realise that it is quite easy to make two straight lines parallel or coincident. *We can, however, lower the error practically to zero if we denote the direction N, after the plate has been put in the spectrograph, by leading the pencil along the sharp edge of the knife T.*

As for the determination of the horizontal deviation  $\gamma$ , an error of  $\psi = 1^\circ$  in adjustment does not play any appreciable part in the determination of  $\gamma$ , being in that case smaller than  $1'$ .

Should it be necessary to determine both  $\gamma$  and  $\epsilon$  with an error smaller than  $1'$ , the method described in the paragraph 2 (a) or 2 (b) can be used, not only for the determination of  $\gamma$ , but also for  $\epsilon$ . For this purpose it is only necessary to put the plate in the spectrograph in such a way that the direction M is parallel with the sharp edge of the knife T, and to use for the evaluation of  $\epsilon$  the equations (3) or (5), adapted in the corresponding way.

##### 5. Grinding the Plate with the Prescribed Angle between the Cut Plane D of the Plate and the Lattice Plane S.

If it is necessary to cut the plate from the crystal in such a way that the angle between the cut plane D of the plate and the lattice plane S has the prescribed value  $\delta_0$ , then—after the plate is cut and after the cut surface is ground—we determine the horizontal deviation  $\gamma$  and the vertical deviation  $\epsilon$ , as described in the paragraph 2 or 3. Using the

equations (11) and (12) we determine the deviation  $\phi$  of the line of intersection P from the direction N and the angle  $\delta$  between the planes D and S. The angle  $\delta$  lies in the plane perpendicular to the line of intersection P. Knowing the angles  $\phi$  and  $\delta$ , we denote the direction of the line of intersection P on the plate and we grind the plate perpendicularly to the line of intersection P by an angle  $\Delta\delta = |\delta - \delta_0|$  to the prescribed angle  $\delta_0$ .

The advantage of the described method consists in the possibility of controlling the cut plane after grinding the plate and to find out how successful we were in the grinding of the plate. It is always possible to remove inaccuracies in grinding by a supplementary correction.

### 6. The Application to Piezoelectricity.

In technical practice it is necessary that the piezoelectric oscillators do not change the frequency of their own vibrations with the temperature, *i. e.* that they have, as we say, the smallest temperature coefficient<sup>(9), (10), (11), (12)</sup>. This is the primary condition for the manufacturing of oscillators in the form of plates and rods. The temperature coefficient depends on how the plates or rods are cut from the crystal. For example, the plates for thickness vibrations possessing the minimum temperature coefficient have to be cut from a quartz crystal, as shown in figs. 12 or 13. The system of co-ordinates in a quartz crystal is chosen in agreement with Voigt<sup>(8)</sup> in such a way that the Y-axis goes out from the crystal under one of the faces R and the X-axis coincides with one of the polar axes.

Fig. 12.

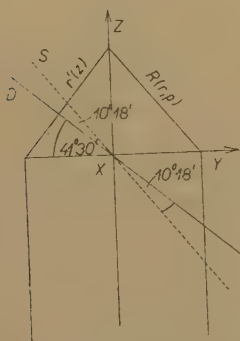
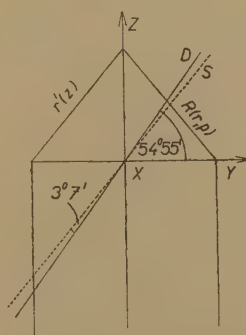


Fig. 13.



In the case shown in fig. 12 the cut plane D of the plate is parallel with the X-axis and has, from the face R, the deviation of  $10^\circ 18'$ . In the case shown in fig. 13 the cut plane D of the plate is also parallel with the X-axis and has, from the surface  $r'$ , the deviation of  $3^\circ 7'$ . Since the Bragg angle for the  $K\alpha_1$  line of copper is equal to  $\vartheta = 13^\circ 19'$ , both for the R and  $r'$  faces, it is possible for measurement of plates cut from a quartz crystal, as shown in figs. 12 or 13, to seek for the determination of  $\gamma$  and  $\epsilon$  when the lattice planes R and  $r'$  are taken as the reflection planes S. The characteristic direction N for the adjustment of the plates in the spectrograph is in both cases the polar or electric axis X, which coincides with the line of intersection P of the cut plane D and lattice plane R or  $r'$ .

7. *Summary.*

The advantage of the described method consists above all in the fact that the determination of both deviations  $\gamma$  and  $\epsilon$  is possible by a single placing of the plate in the spectrograph and by one exposure, with a precision which is sufficient for many technical problems. On the basis of the measured deviations  $\gamma$  and  $\epsilon$ , it is possible to make a correction to the cut plane D by grinding the plate according to the paragraph 5.

A further advantage is to be considered in the possibility of the control of the plate after grinding. This control enables us to find out if the correction of the cut plane was correctly done.

Finally, one should mention that all the equations necessary for the determination of the deviations  $\gamma$  and  $\epsilon$  were derived only for the beam coming from the source Z, going through the scratch V and falling on the film F. It would be necessary, especially for the determination of large deviations  $\epsilon$ , to give a precise theory of representation in Seemann's spectrograph provided with a scratch V.

These investigations were performed during the war in the laboratories of Vojenské telegrafní dílny at Kbely, near Prague, and were finished in the year 1942 but not published. The authors only submit them now by permission of the Factory Workers' Trade Union, to which they extend many thanks for the permission to publish them.

*References.*

- (1) А. В. Шубников, Труды лаборатории кристаллографии, Выпуск 2, Издательство Академии Наук СССР, 1940 год.  
A. V. Shubnikov, Reports from the Crystallographic Laboratory, tome ii. published by the Academy of Science of U.S.S.R. (1940).
- (2) 'Electronics,' November Nr. (1940).
- (3) G. K. Burns, 'Bell System Technical Journal,' t. xix. p. 516 (1940).
- (4) В. И. Архаров, Журнал технической физики, т. x. стр. 1268 (1940 г.)  
V. I. Arkharov, 'Journal of Technical Physics,' t. x. p. 1268 (1940).
- (5) В. И. Архаров и В. С. Аверкиев, Журнал технической физики, т. x. стр. 1276 (1940 г.).  
V. I. Arkharov and V. S. Averkiev, 'Journal of Technical Physics,' t. x. p. 1276 (1940).
- (6) M. Majima and S. Togino, Scientific Papers of the Institute of Physical and Chemical Research, t. vii. pp. 75 and 259 (1927).
- (7) H. Seemann, *Annalen der Physik*, t. li. p. 391 (1916); t. liii. p. 461 (1917); t. lvi. pp. 1 and 793 (1930).
- (8) W. Voigt, *Lehrbuch der Kristallphysik*, Leipzig, p. 702 (1910 and 1928).  
'Notation for Piezo-electric Quartz,' The National Physical Laboratory, London (1938).
- (9) W. G. Cady, 'Piezoelectricity,' McGraw-Hill, New York and London, 1946.
- (10) P. Vigoreux, 'Quartz Oscillators and Resonators,' H.M. Stationery Office, London. 1931.
- (11) W. L. Bond and E. J. Armstrong, "Use of X-Rays for Determining the Orientation of Quartz Crystals," The Bell System Technical Journal, October 1943.

Prague.



XL. *The Propulsive Force produced by an Intermittent Jet of Gas.*

By L. J. KASTNER, M.A.\*

[Received August 2, 1945.]

THE problem of the propulsion of aircraft by means of high-velocity jets of gas is now known to have been solved successfully, and it appears that the type of engine most generally favoured for this type of propulsion consists basically of a combination of gas turbine and turbo-compressor, the compressor supplying air to burn oil substantially at constant pressure, the hot gases of combustion giving up part of their energy in the turbine—the power output of which is used to drive the compressor—and then passing on to expand in a nozzle. The high velocity jet so formed is used to propel the aircraft by direct reaction. When the aircraft is in level flight at constant speed, it would seem that both the mass of gas per unit time leaving the propelling nozzles and the exit velocity of this gas should be constant, and that the product of these two quantities should equal the propulsive force.

It is, however, not so widely known that when a high-speed aircraft is driven by the more conventional type of power unit comprising a reciprocating engine driving an airscrew, a useful increase in effective power and speed can be obtained by making use of the kinetic energy in the exhaust gases from the engine cylinders to provide an auxiliary reactive force, additional to that produced by the airscrew. The magnitude of the reactive force produced in this case is more difficult to estimate than that referred to in the previous paragraph, because the exhaust process of the reciprocating engine is not a continuous one. Thus, in considering the discharge from any one cylinder, neither the mass per second discharged nor the velocity of discharge can be assumed to be constant. The exhaust jet is therefore intermittent and the instantaneous propulsive force also varies with time. As far as the writer is aware, no theoretical treatment of the problem is generally available, although the practical value of the auxiliary exhaust jets is not in doubt. For this reason an attempt to provide a simple solution may have some interest.

Consider a body of mass  $m + \Delta m$  moving initially with velocity  $v$  and suppose the body ejects a mass  $\Delta m$  in the direction opposite to that of the original direction of motion. After the separation, suppose that the large body has a velocity  $V$  and let the relative velocity between the two bodies be  $u$ . Then, by the principle of conservation of momentum,

$$(m + \Delta m)v = mV + \Delta m(V - u)$$

or

$$V = v + \frac{\Delta m \cdot u}{(m + \Delta m)}.$$

\* Communicated by the Author.

Assuming  $\Delta m$  to be small compared with  $m$ ,

$$V - v = \frac{\Delta m}{m} \cdot u.$$

Suppose  $\Delta m$  represents a mass of gas of density  $\rho$ , ejected from an exhaust orifice of area  $A$  in time  $\Delta t$ . Then

$$\Delta m = \rho \cdot A \cdot u \cdot \Delta t,$$

and, calling

$$V - v = \Delta v,$$

we have

$$\Delta v = \frac{\rho A u^2}{m} \cdot \Delta t,$$

or the propulsive force

$$m \frac{\Delta v}{\Delta t} = \rho A u^2 = u \frac{\Delta m}{\Delta t} \cdot \dots \dots \dots (i)$$

In the case of an aircraft fitted with a reciprocating engine, both  $u$ , the gas velocity through the exhaust orifice, and

$$\frac{\Delta m}{\Delta t},$$

the rate of mass-discharge of the exhaust gases, vary with time, and both must be evaluated before the thrust can be estimated. It may be noted here that even a small reactive thrust can deliver useful power, if the aircraft is travelling at high speed.

Suppose figs. 1(A) and 1(B) refer respectively to the fall of pressure during the exhaust period and to the available escape area for one cylinder of a reciprocating engine. Then, referring to fig. 1(A), over an interval of time  $\Delta t$ , the mean cylinder pressure is

$$P_m = P_1 - \frac{\Delta P}{2},$$

and, if  $\rho_m$  is the density corresponding to  $P_m$ , then the critical velocity which determines the mass of gas flowing through the orifice per unit time is

$$u_t = \sqrt{2g \frac{\gamma}{\gamma+1} \cdot \frac{P_m}{\rho_m}}, \quad \dots \dots \dots (ii)$$

assuming adiabatic expansion in the orifice and taking  $\gamma$  to be the adiabatic exponent. From fig. 1(B), the mean escape area of the orifice during the interval  $\Delta t$  is

$$A_m = A_1 + \frac{\Delta A}{2}.$$

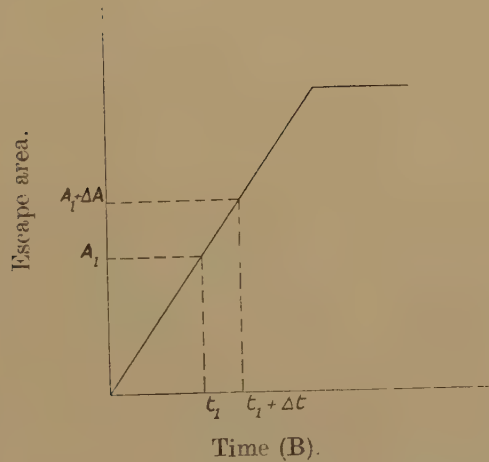
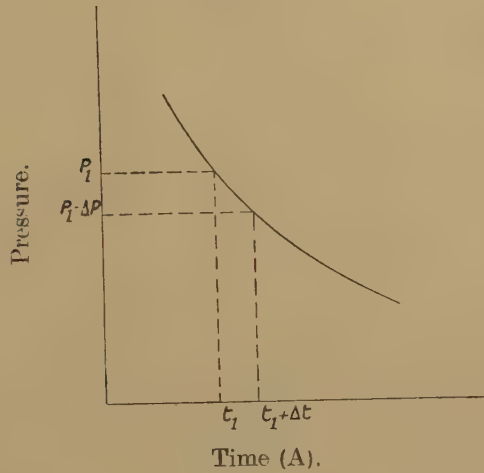
Thus, in time  $\Delta t$  the mass discharged per second will be given by the expression

$$\frac{\Delta m}{\Delta t} = \rho_t u_t A_m \cdot \dots \dots \dots (iii)$$

Here  $\rho_t$  refers to the density of the gas at the throat of the orifice and is known, since

$$\frac{\rho_t}{\rho_m} = \left( \frac{2}{\gamma+1} \right)^{\frac{1}{\gamma-1}}.$$

Fig. 1.



So far it has been assumed that the ratio between the exhaust back pressure and the cylinder pressure is less than the critical value, and for the major part of the discharge this will be so, especially when the aircraft is flying at considerable altitude in a rarefied atmosphere. Towards the end of the discharge period, however, the pressure in the cylinder will have fallen so far that the ratio between the back pressure (atmospheric) and the cylinder pressure will now be greater than the

critical value, so that the discharge will be dependent on the back pressure  $P_x$ . The velocity of flow is then given by the equation

$$u_x = \sqrt{2g \frac{\gamma}{\gamma-1} \cdot \frac{P_m}{\rho_m} \left( 1 - \left[ \frac{P_x}{P_m} \right]^{\frac{\gamma-1}{\gamma}} \right)} \quad \dots \quad (iv)$$

The corresponding density

$$\rho_x = \rho_m \left( \frac{P_x}{P_m} \right)^{1/\gamma},$$

and the mass discharged per second

$$\frac{\Delta m}{\Delta t} = \rho_x u_x A_m', \quad \dots \quad (v)$$

where  $A_m'$  is the appropriate value of the escape area.

The solution of the equations (iii) and (v) will allow the variation of

$$\frac{\Delta m}{\Delta t}$$

with time to be found, as a first step towards the solution of equation (i).

The velocity  $u$  of equation (i) represents the velocity with which the gas leaves the discharge orifice, and in the ideal case this corresponds to the velocity with which it is ejected from the cylinder. In practice, however, the propulsive nozzle must be some distance from the cylinder and there will be some loss of energy. For this reason it has been considered advisable to suppose that the effective value of the discharge velocity never exceeds the value of  $u_t$  given by equation (ii), which equation is therefore assumed to apply until the cylinder pressure falls so far that it becomes necessary to use equation (iv) to calculate the velocity. After this stage has been reached, values obtained by the use of the latter equation are taken to be the true velocities of flow.

#### *Application to a Particular Case.*

The calculation of the thrust due to the exhaust gases is best illustrated with reference to a particular example.

Consider the case of a reciprocating engine, the mechanical power output of which is used to drive the airscrew of a high-speed aircraft. It is desired to find what proportion of the total effective thrust on the machine is contributed by the jets of exhaust gas leaving the engine cylinders, supposing these jets to be utilized for this purpose. The following data are necessary for a numerical solution of the problem and are in accord with modern practice :—

Mean forward speed of aircraft=400 miles per hour.

Altitude=16,000 ft., atmospheric pressure=8 lb. per sq. in. (abs.) approx.

Power output at engine crankshaft=1600 h.p.

Engine dimensions=14 cylinders, swept volume per cylinder 186 cubic in., rotational speed 2400 r.p.m.



The mass of air taken in per cylinder per cycle is given by

$$M = \eta \cdot \frac{PV}{RT},$$

where P and T refer respectively to the pressure and temperature of the charge before it enters the cylinder, V is the swept volume of the cylinder, and  $\eta$  is the volumetric efficiency. The engine will be supposed to be "supercharged" to give ground-level atmospheric pressure at the stated altitude, so that  $P = 14.7$  lb. per sq. in. (abs.), and the value of T might be  $580^\circ$  F. abs., to allow for heating in the supercharger.  $\eta = 0.8$  would be a reasonable assumption.

$$\text{Thus } M = \frac{0.8 \times 14.7 \times 186}{53.2 \times 580 \times 12} = 0.0059 \text{ lb.}$$

If the air-fuel ratio were 13 : 1 and if the mass of gas in the cylinder clearance space equalled 5 per cent. of the mass taken in, the total mass of gas in the cylinder would be

$$\left(1 + \frac{1}{13}\right)(0.0059) \frac{21}{20} = 0.0067 \text{ lb.}$$

To find the density of the gas in the cylinder at the moment when the exhaust valve first begins to open, the effective volume must be known. This is less than the whole cylinder volume because the exhaust valve begins to open before the piston reaches the bottom dead centre position. It is a fair assumption, with petrol engines of normal valve timing and compression ratio, to suppose that the cylinder volume at the instant considered equals the swept volume, namely 186 cubic in. in our case. The density of the cylinder gas when the exhaust valve begins to open is therefore

$$\rho_1 = \frac{0.0067 \times 1728}{186} = 0.062 \text{ lb. per cubic ft.}$$

The only additional information required for a solution of the problem concerns the effective area of the exhaust orifice, and fig. 2 shows this quantity related to time, for the engine speed considered. The diagram has been somewhat simplified in form for the present purpose, and allowance has been made for a mean constant coefficient of discharge of value 0.7.

#### *Numerical Calculation of the Quantity $u \frac{\Delta m}{\Delta t}$ .*

From equations (ii) and (iii) an expression for the rate of discharge can be derived, having the form

$$\frac{\Delta m}{\Delta t} = A_m \left( \frac{2}{\gamma + 1} \right)^{\frac{1}{\gamma - 1}} \sqrt{2g \frac{\gamma}{\gamma + 1} P_m \rho_m} \quad \therefore \quad \text{(vi)}$$

This may be solved if the variation with time of both the discharge area and the cylinder pressure are known. As yet, the variation with time of the latter quantity is not known.

Suppose the cylinder pressure to fall from its initial value  $P_1$  to a smaller value  $P_2$ , consequent on the opening of the exhaust orifice. The mass in the cylinder corresponding to pressure  $P_1$  was

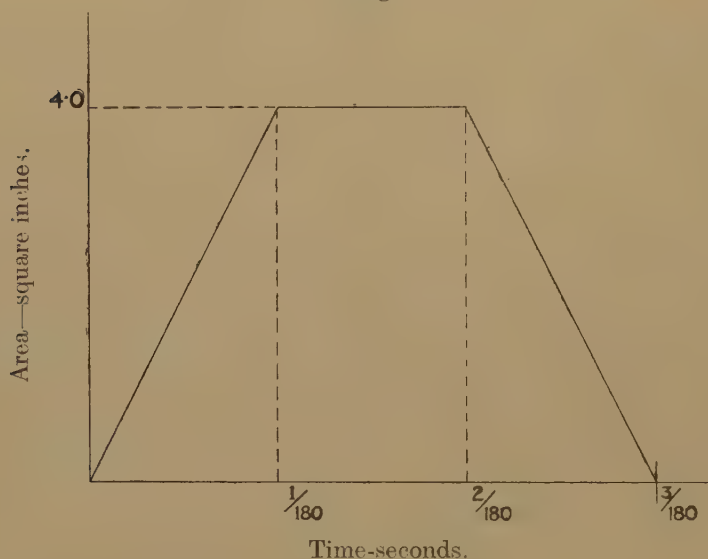
$$M_1 = \frac{P_1 V}{RT_1},$$

where  $T_1$  refers to the temperature and  $V$  to the cylinder volume. At pressure  $P_2$  the mass remaining in the cylinder is

$$M_2 = \frac{P_2 V}{RT_2},$$

assuming the cylinder volume to be substantially unchanged.

Fig. 2.



Thus, the mass ejected

$$M_1 - M_2 = \frac{V}{R} \left( \frac{P_1}{T_1} - \frac{P_2}{T_2} \right),$$

and the proportion of the original mass ejected

$$\frac{M_1 - M_2}{M_1} = 1 - \frac{P_2 T_1}{P_1 T_2},$$

which becomes, if the gas remaining in the cylinder be supposed to expand adiabatically,

$$\frac{M_1 - M_2}{M_1} = 1 - \left( \frac{P_2}{P_1} \right)^{1/\gamma}.$$

Supposing  $P_1 - P_2$  to be small, so that  $\frac{1}{2}(P_1 + P_2) = P_m$ , and calling  $M_1 - M_2 = \Delta m$ , this can be written in the form

$$\frac{\Delta m}{M_1} = 1 - \left( \frac{P_2}{P_1} \right)^{1/\gamma} \quad \dots \dots \dots \text{(vii)}$$

Equations (vi) and (vii) combined now give

$$\Delta t \cdot A_m \left( \frac{2}{\gamma+1} \right)^{\frac{1}{\gamma-1}} \sqrt{2g \frac{\gamma}{\gamma+1} P_m \rho_m} = M_1 \left[ 1 - \left( \frac{P_2}{P_1} \right)^{1/\gamma} \right]. \quad (\text{viii})$$

This equation, solved in a series of steps, enables both a "pressure-time" relation and a "mass discharged-time" relation to be found, and so permits

$$\frac{\Delta m}{\Delta t}$$

to be evaluated.

If  $t_1$  refers to the time at the end of a small interval immediately following the instant at which the exhaust orifice first begins to open, equation (viii) becomes, for this interval,

$$t_1 \cdot K \frac{t_1}{2} \cdot \left( \frac{2}{\gamma+1} \right)^{\frac{1}{\gamma-1}} \sqrt{2g \frac{\gamma}{\gamma+1} P_m \rho_m} = M_1 \left[ 1 - \left( \frac{P_2}{P_1} \right)^{1/\gamma} \right],$$

or 
$$\frac{1}{2} C \sqrt{P_m \rho_m} = M_1 \left[ 1 - \left( \frac{P_2}{P_1} \right)^{1/\gamma} \right],$$

where  $K$  and  $C$  are constants. This enables  $t_1$  to be found for a selected small pressure difference  $P_1 - P_2$ . If now the pressure falls from  $P_2$  to  $P_3$  during a second small interval  $t_2 - t_1$ , then

$$\left( \frac{t_2^2 - t_1^2}{2} \right) C \sqrt{P_m \rho_m} = M_1 \left\{ \left( \frac{P_2}{P_1} \right)^{1/\gamma} - \left( \frac{P_3}{P_1} \right)^{1/\gamma} \right\},$$

and so on,  $P_m$  here being  $\frac{P_2 + P_3}{2}$ . When the pressure ratio  $\frac{P_x}{P_m}$  becomes greater than the critical value, namely,

$$> \left( \frac{2}{\gamma+1} \right)^{\frac{\gamma}{\gamma-1}},$$

the term 
$$\left( \frac{2}{\gamma+1} \right)^{\frac{1}{\gamma-1}} \sqrt{2g \frac{\gamma}{\gamma+1} P_m \rho_m}$$

on the left-hand side of equation (viii) becomes

$$\sqrt{2g \frac{\gamma}{\gamma+1} P_m \rho_m \left[ \left( \frac{P_x}{P_m} \right)^{2/\gamma} - \left( \frac{P_x}{P_m} \right)^{\frac{\gamma+1}{\gamma}} \right]},$$

but this only applies to the later stages of the process. Between

$$t = \frac{1}{180} \text{ sec. and } t = \frac{2}{180} \text{ sec.}$$

the escape area  $A_m$  is constant, as shown in fig. 2.

The results of the calculation are given in Table I., based on an initial cylinder pressure of 105 lb. per sq. in. (abs.) and an initial mass, as obtained above, of 0.0067 lb. A value of  $\gamma = 1.3$  has been adopted for the hot gases.

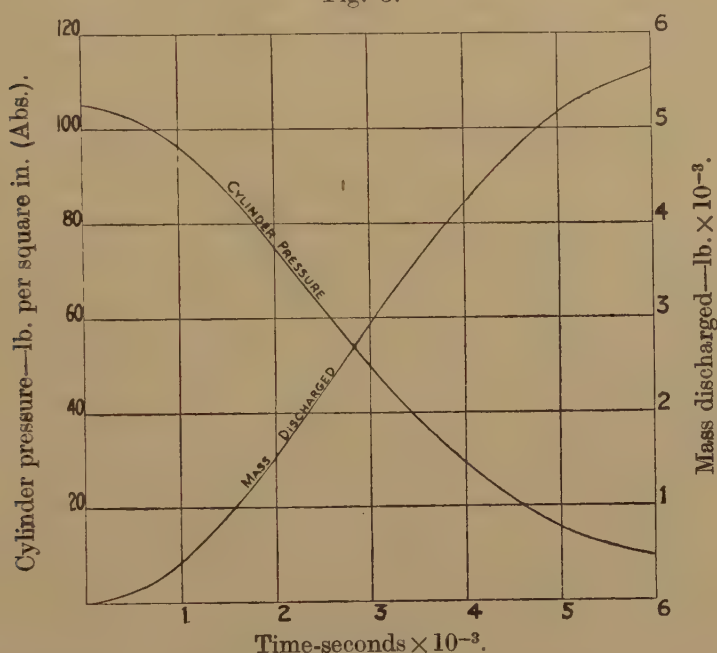
TABLE I.

	Initial Pressure lb./sq. in. (abs.)	Final Pressure lb./sq. in. (abs.)	Mean Pressure lb./sq. in. (abs.)	Time-seconds	Mass discharged (lb.)	$\frac{\Delta m}{\Delta t}$ lb. per sec.	Force $\frac{u}{g} \frac{\Delta m}{\Delta t}$ lb.
1st interval	105	100	102.5	$0.756 \times 10^{-3}$	$0.246 \times 10^{-3}$	0.57	52.6
2nd	100	95	97.5	$1.073 \times 10^{-3}$	$0.496 \times 10^{-3}$	0.875	80.2
3rd	95	90	92.5	$1.34 \times 10^{-3}$	$0.75 \times 10^{-3}$	1.043	95.2
4th	90	85	87.5	$1.575 \times 10^{-3}$	$1.008 \times 10^{-3}$	1.177	106.4
5th	85	80	82.5	$1.79 \times 10^{-3}$	$1.262 \times 10^{-3}$	1.27	114.1
6th	80	75	77.5	$1.997 \times 10^{-3}$	$1.531 \times 10^{-3}$	1.34	119.6
7th	75	70	72.5	$2.199 \times 10^{-3}$	$1.80 \times 10^{-3}$	1.385	123.0
8th	70	65	67.5	$2.398 \times 10^{-3}$	$2.07 \times 10^{-3}$	1.42	125.0
9th	65	60	62.5	$2.577 \times 10^{-3}$	$2.345 \times 10^{-3}$	1.425	124.1
10th	60	55	57.5	$2.781 \times 10^{-3}$	$2.63 \times 10^{-3}$	1.41	122.0
11th	55	50	52.5	$2.991 \times 10^{-3}$	$2.92 \times 10^{-3}$	1.38	118.0
12th	50	45	47.5	$3.209 \times 10^{-3}$	$3.202 \times 10^{-3}$	1.338	113.0
13th	45	40	42.5	$3.439 \times 10^{-3}$	$3.51 \times 10^{-3}$	1.292	108.0
14th	40	35	37.5	$3.686 \times 10^{-3}$	$3.822 \times 10^{-3}$	1.24	102.0
15th	35	30	32.5	$3.956 \times 10^{-3}$	$4.147 \times 10^{-3}$	1.165	94.3
16th	30	25	27.5	$4.257 \times 10^{-3}$	$4.48 \times 10^{-3}$	1.075	85.4
17th	25	20	22.5	$4.607 \times 10^{-3}$	$4.84 \times 10^{-3}$	0.945	73.2
18th	20	15	17.5	$5.033 \times 10^{-3}$	$5.2 \times 10^{-3}$	0.725	54.6
19th	15	10	12.5	$5.89 \times 10^{-3}$	$5.602 \times 10^{-3}$	0.365	22.3



The curves of fig. 3 represent the calculated variation with time of both cylinder pressure and gas discharge, and from the slope of the "gas discharge-time" curve, the value of  $\frac{\Delta m}{\Delta t}$  may be found at any instant. The dotted curve of fig. 4 shows the variation of this quantity, and the lower, full-line curve of the same figure represents the variation of  $\frac{u \Delta m}{g \Delta t}$ , the impulsive force in lb. due to the jet. The area bounded by the  $\frac{u \Delta m}{g \Delta t}$  curve and the horizontal axis is found by graphical integration to

Fig. 3.



Curves relating "mass discharged" and "cylinder pressure" to "time."

equal, in the present case, 0.483 pound-seconds, and the mean force exerted by the jet will be given by this quantity divided by the time taken to complet one cycle\* at the specified speed, namely, 2400 r.p.m. Thus the mean propulsive force due to the jet is

$$\frac{0.483}{1/20} = 9.66 \text{ lb. per cylinder.}$$

The thrust h.p. produced by the exhaust jets of 14 cylinders at an airspeed of 400 miles per hour would then be

$$\frac{400 \times 88 \times 14 \times 9.66}{550 \times 60} = 144 \text{ h.p. approximately.}$$

Remembering that the power output at the engine crankshaft was

---

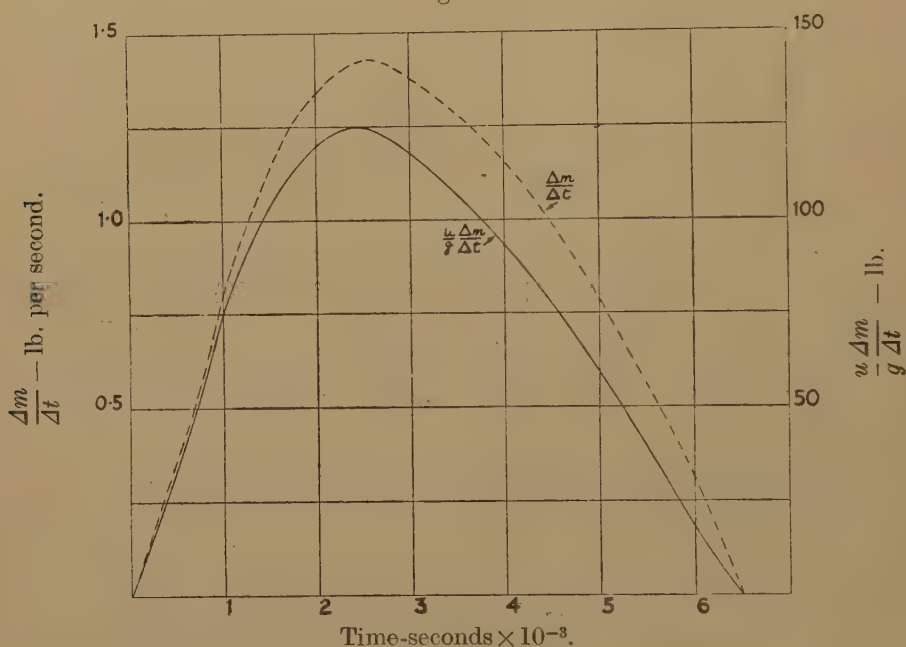
\* Two revolutions per cycle with a four-stroke engine.

1600 h.p. and assuming an airscrew efficiency of 80 per cent., the effective thrust h.p. at the airscrew  $= 0.8 \times 1600 = 1280$  h.p. Hence the gain due to the jets

$$= \frac{144}{1280} = 11.3 \text{ per cent.}$$

It will be noted from fig. 4 that the time during which the thrust is effective is only about  $6.5 \times 10^{-3}$  seconds per cylinder per cycle in the present case, whilst fig. 2 shows that the exhaust orifice is open for over twice as long as this. It is clear, however, that by far the greater proportion of the original contents of the cylinder is expelled during

Fig. 4.



Curves relating " $\frac{\Delta m}{\Delta t}$ " and " $\frac{u \Delta m}{g \Delta t}$ " to "time."

the early part of the exhaust period, and that, although the engine piston is aiding the process of expulsion during the latter part of the period, the mass to be ejected is now too small and the cylinder pressure too low for any appreciable exhaust thrust to be developed. For this reason, although it is realized that the treatment given above is approximate only, it is not considered that any serious error is involved in neglecting the motion of the piston. This can be taken into account, if desired, but only at the cost of greatly complicating the analysis.

The final result of the analysis, in its present form, gives, the writer believes, a result in good general agreement with practical experience, but, in view of the uncertainty existing regarding the various losses which would occur in a real case and because of the scarcity of experimental data, no special accuracy is claimed for it.

XLI. *Kinematic Relativity.*

By Professor W. WILSON, F.R.S.\*

[Received March 1, 1946.]

IT was not my intention, in my short paper† on this subject, to deny, or seem to deny, the possibility of *contemplating* free particles other than Professor Milne's fundamental particles, or to deny to him the same freedom of inquiry which Newton and Einstein enjoyed. I have, of course, a natural bias in favour of the Einstein-Minkowski type of relativity; but I am trying to avoid being enslaved by it and I can make the concession to Milne and his school that, his  $\mathbf{V}t$  and  $\mathbf{P}-\mathbf{V}t$ , in the formula

$$\frac{1}{\mathbf{Y}} \frac{d\mathbf{V}}{dt} = \frac{\mathbf{P}-\mathbf{V}t}{\mathbf{X}} G(\xi), \quad . . . . . (1)$$

are not vectors and that the premises of kinematic relativity seem to me to imply this. His theory, therefore, does not now appear to me to suffer from the particular weakness or defect which I suspected. The most important point at issue between us is consequently settled in favour of Milne's theory. I am giving below a new and very much shorter derivation of the formula (1), which I call Milne's fundamental formula; but before proceeding to it I wish to refer briefly to the arguments in the recent papers on this subject, more especially Milne's.

I would begin by drawing attention to a sentence in my paper‡ which Professor Milne seems to have overlooked, namely, "I agree that  $\mathbf{X}$  is an invariant and that  $\mathbf{P}-\mathbf{V}t$  is the space part of a 4-vector." The sentence indicates, what is in fact the case, that the suggestion that  $\mathbf{P}-\mathbf{V}t$  is a vector come from Milne or from some exponent of his theory. I received this impression from his 'Relativity, Gravitation and World Structure,' more especially from page 99, where will be found the statement: "Thus in Fig. 11, the acceleration of the  $\mathbf{V}$ -particle at  $P$  at time  $t$  is along the *vector* [my italics]  $P_0P=\mathbf{P}-\mathbf{V}t$ ." I give the quotation to show how I got the notion that  $\mathbf{P}-\mathbf{V}t$  is a vector, and not as an argument. Then, of course,  $\mathbf{P}$  is a vector, *i. e.* the space part of a 4-vector, in any case, and  $\mathbf{V}t$  does in fact appear in the context  $\mathbf{P}-\mathbf{V}t$ , and it is not at all so very obvious how it can be a vector in the private Euclidean space of an observer, as Dr. Whitrow would express it, and yet not the space part of a 4-vector, while  $\mathbf{P}$  from which it is subtracted is both a vector in the observer's private space and also the space part of a 4-vector. The long arguments of Milne, Whitrow and Newing do not seem to me to render any adequate elucidation of this important point.

---

\* Communicated by the Author.

† Phil. Mag. ser. 7, vol. xxxv. p. 241 (April 1944).

‡ Loc. cit. p. 243.

Nevertheless I believe they are right in this contention and have discussed it in my way in the derivation given below.

The assumption I made was, in effect, that  $\mathbf{V}t$  was identical with  $\mathbf{P}_0$ , the displacement from O of the fundamental particle  $\mathbf{P}_0$ , which has the same velocity as the free particle at P and that  $\mathbf{P}_0$  is the space part of a 4-vector. This assumption leads of necessity to the vanishing of  $\mathbf{P}-\mathbf{V}t$ . I did not assume, as Dr. Whitrow puts it \* “that  $\mathbf{V}t$  is the space part of a 4-vector *in order to prove* [my italics] that the motion of *any* free particle reduces to that of a fundamental particle.” The result follows from my premises and I will not have it that it is a *petitio principii*. Professor Milne † stigmatises an identity of mine, numbered (20) in his paper, as false; but he does agree that the *components* on the two sides of (20) along  $\mathbf{V}$  are equal. But surely the plain meaning of the equation is obvious. It is sufficient for the purposes I had in mind, namely the derivation of Milne’s

$$\frac{1}{Y^{1/2}} \frac{d}{dt} \left( \frac{\mathbf{V}}{Y^{1/2}} \right) = - \frac{1}{X} \left( \mathbf{P} - \mathbf{V} \frac{Z}{Y} \right), \dots \dots \dots (2)$$

among other things, to deal with the special case (to which I shall refer again under this name) where  $\mathbf{V}$  is along the line OP joining the observer to the free particle at P, in one direction or the opposite. It was not (20) which led to the vanishing of  $\mathbf{P}-\mathbf{V}t$ ; but something quite different. It does, in fact, lead to (2).

The special case just mentioned is important because, when it holds, Milne’s cosmological principle requires that the acceleration of the free particle shall be along the line OP (in one direction or the opposite), and because, for any free particle, there are observers so placed that the free particle is moving (in one direction or the opposite) along the line joining the observer and the free particle.

Similar observations apply to Milne’s criticism of my enunciation of the Lorentz transformation as

$$\mathbf{D} = \frac{\mathbf{D}' + \mathbf{U}t'}{(1 - \mathbf{U}^2/c^2)^{1/2}},$$

which, as the context indicates, applies to the special case where  $\mathbf{D}$ ,  $\mathbf{D}'$  and  $\mathbf{U}$  are in the same direction. I do not need to elaborate this any further, except just to remark that the usual form of the Lorentz transformation, *e. g.*

$$\left. \begin{aligned} x' &= \gamma(x - vt), \\ y' &= y, \\ z' &= z, \\ t' &= \gamma \left( t - \frac{vx}{c^2} \right) \end{aligned} \right\}, \dots \dots \dots (3)$$

\* Phil. Mag. ser. 7, vol. xxxvi. p. 173 (March 1945).

† Phil. Mag. ser. 7, vol. xxxvi. p. 134 (February 1945). There is a misprint

in equation (20):  $\frac{1}{Y}$  on the right-hand side should be replaced by  $\frac{1}{Y^{1/2}}$ .





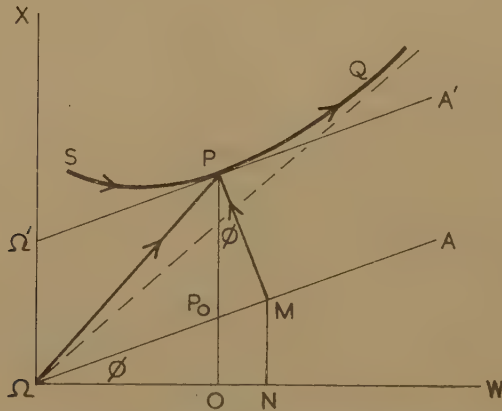
represented in the figure is the particular  $\phi$  of the transformation from an observer O (special case) to that observer in whose system the free particle is momentarily at rest (new W axis along  $\Omega A$ ). I wish to anticipate any suggestion of mysticism or of building on geometrical intuition. The figure is neither more nor less than a concise and very useful symbolization of the Lorentz transformation.

We can make use of the familiar formula for the curvature of a plane curve, namely

$$\frac{\frac{d^2x}{dw^2}}{\left\{1 + \left(\frac{dx}{dw}\right)^2\right\}^{3/2}} = \frac{1}{R}^* \dots \dots \dots (6)$$

Remembering that  $w = \sqrt{-1}ct$ , we may give (6) the form

$$-\frac{1}{c^2} \frac{dV}{dt} = \frac{1}{R}, \dots \dots \dots (6a)$$
$$\left(1 - \frac{V^2}{c^2}\right)^{3/2}$$



where  $x$  has been identified with  $P$ , the  $x$  co-ordinate of the free particle.

Now we inquire what we are given, if anything, which determines the curvature. For any equivalent observer one free particle is distinguished from another (a) by its co-ordinates  $\mathbf{P}$  and  $\mathbf{W}$ , i. e. by the vector  $\mathbf{S} = \vec{\Omega P}$  (see figure) whose absolute value is

$$\Omega P = \sqrt{OP^2 + \Omega O^2} = \sqrt{P^2 + W^2},$$

and (b) by  $\mathbf{V}$ , or by the angle  $\phi$ , since  $V/\sqrt{-1}c = \tan \phi$ ; or finally by the vector  $\pi$ , whose absolute value is equal to  $MP$  (see figure), the perpendicular distance between  $\Omega A$  and  $\Omega' A'$ . There appears to be no other

possibility except 
$$\frac{1}{R} = \frac{K}{S^2} \pi, \dots \dots \dots (7)$$

\* The square roots in this and other formulæ may be taken as positive and any ambiguity may be relegated to the invariant  $K$  which appears below.



Whether mystical or not, the use of  $\sqrt{-1}$  spares us much laborious effort.

The nature of the vector, or, I should say, the pseudo-vector,

$$\mathbf{P}_0 = \mathbf{V}t,$$

is made quite clear by the figure—especially if we consider the special case. The Lorentz transformation is equivalent, as we have seen, to a rotation of the  $x, w$  axes through some angle  $\phi$ , not necessarily the particular  $\phi$  mentioned above. The reader can construct for himself in the figure new axes  $x', w'$  with the origin  $\Omega$ , so that the angle between  $w$  and  $w'$  is, say,  $\phi'$ . In this new system, which is that of another equivalent observer, the  $x'$  co-ordinate of the free particle will be  $O'P$  which is perpendicular to the  $w'$  axis, and  $O'P$  will cut the invariant line  $\Omega A$  at some point  $P_0'$ , not coincident with  $P_0$ , so that

$$O'P_0' = \mathbf{V}'t'$$

is derived by a method or process totally different from that represented by the transformation (5) above.

I am, of course, aware that there are limits to the possible values of the angle  $\phi$ , or, since it is not a real angle, shall I say its cosine ranges from unity to infinity. The angle  $x\Omega w$  in the figure includes angles whose cosines are purely imaginary. They are not quite devoid of physical significance; but what physical significance they have is outside the scope of this paper.

The derivation of Milne's fundamental equation (1) or (2) above, is an anticipation (since I did it before his last paper appeared) of his suggestion that I should realize wherein the newness of his theory lies if I were to seek "to enrich the theory by some new type of kinematic argument." The argument given here has at least a new form, and I am inclined to believe that I have at last grasped the significance of kinematic relativity—and may I add here that I am much indebted to the contributors of the three replies in the 'Philosophical Magazine' and also to my correspondents, T. Lewis of Aberystwyth and Dr. Alfred Schild of Toronto.

## XLII. *Forced Vibrations of a Whirling Wire* \*.

By A. VICTOR MASKET, Ph.D.,

Mechanics and Electricity Division, Naval Research Laboratory,  
Washington, D.C. †

[Received July 5, 1945.]

The method of the Laplace transformation and contour integration is applied to obtain a solution for the partial differential equation governing the transverse vibrations of a long thin wire subject to a

\* Published by permission of the U.S. Navy Department. The opinions expressed in this article are those of the writer and do not necessarily represent those of the U.S. Navy.

† Communicated by the Author,



combined static tension and centrifugal force when the moving end is forced to execute a sinusoidal motion. The novel features of the solution are that the eigen functions are in general transcendental Legendre functions, which are not orthogonal over any interval of the defined range of validity.

### Statement of Problem.

The solution for the classical problem of the transverse vibrations of a whirling chain in terms of an infinite set of orthogonal polynomials of odd order, is a familiar boundary value problem in partial differential equations\*, † cited in the literature. An interesting variation of this problem was brought to the writer's attention by Mr. Frank Louckes in the course of some experiments.

A long thin wire of density  $\rho$ , cross-section  $A$ , and elastic modulus  $E$  is stretched down the axis of a closed tube of length  $l$  and fixed at the ends in a state of uniform tension  $T_0$ . The transverse vibrations of the wire are to be determined when the tube is whirled horizontally about an axis perpendicular to its length at one end with uniform angular velocity  $\omega$ , and the free end of the tube undergoes a vertical sinusoidal motion  $B \sin \nu t$  which may be attributed to turbulence.

### Method of Solution.

There are two simple cases which may arise in practice :—

*Case 1.*—The tube is considered virtually inextensible during the rotation in comparison with the wire. We require, first, the state of tension during the steady rotation and, second, the maximum allowable value of  $\omega$  to avoid "buckling" of the wire, that is, the tension shall not fall to zero anywhere.

If  $u(x)$  represents the displacement along the wire which takes place at  $0 \leq x \leq l$ , then during steady rotation the displacement function satisfies the equation

$$E \frac{d^2 u(x)}{dx^2} + \rho \omega^2 x = 0 \quad \dots \dots \dots (1)$$

with end conditions

$$\left. \begin{array}{l} x=0: \quad u(0)=0, \\ x=l: \quad u(l)=T_0 l / EA. \end{array} \right\} \dots \dots \dots (2)$$

The solution is

$$EA u(x) = T_0 x + \frac{1}{6} \rho A \omega^2 (l^2 x - x^3) \quad \dots \dots \dots (3)$$

The state of tension is given by

$$T(x) = EA \frac{du}{dx} = T_0 + \frac{1}{2} \rho A \omega^2 \left( \frac{l^2}{3} - x^2 \right) \quad \dots \dots \dots (4')$$

\* A. G. Webster, 'Partial Differential Equations of Mathematical Physics,' p. 137 (B. G. Teubner, Leipzig, 1933).

† J. C. Slater and N. H. Frank, 'Introduction to Theoretical Physics,' p. 158 (McGraw-Hill Book Company, Inc., New York and London, 1933).

If we require that  $T(x) > 0$  for all  $0 \leq x \leq l$ , then since  $T(x)$  is a minimum for  $x=l$ , the condition on  $\omega$  follows from

$$\frac{1}{3}\rho A\omega^2 l^2 < T_0. \quad (5)$$

*Case 2.*—The tube and wire are of equal elastic modulus. It is well known that in this case

$$T(x) = T_0 + \frac{1}{2}\rho A\omega^2(l^2 - x^2), \quad (4'')$$

and  $\omega$  is limited only by the tensile strength of tube or wire.

We are now in a position to solve the partial differential equation for the transverse vibration of the wire. The equation takes the form

$$\frac{\partial}{\partial x} \left\{ T(x) \frac{\partial v(x, t)}{\partial x} \right\} = \rho A \frac{\partial^2 v(x, t)}{\partial t^2}, \quad (6)$$

where the variable tension  $T(x)$  has the meaning of (4') or (4''), and with boundary conditions

$$t=0: v(x, 0)=0, \partial v/\partial t|_{t=0}=0, 0 \leq x \leq l; \quad (7a)$$

$$x=0: v(0, t)=0; \quad x=l: v(l, t)=B \sin \nu t, t \geq 0. \quad (7b)$$

If we divide (6) by

$$T_0 + \left\{ \frac{1/6}{1/2} \right\} \rho A\omega^2 l^2$$

and make the transformations

$$y^2 = \frac{\frac{1}{2}\rho A\omega^2 x^2}{T_0 + \left\{ \frac{1/6}{1/2} \right\} \rho A\omega^2 l^2}, \quad \tau = \frac{\omega}{\sqrt{2}} t, \quad (8)$$

the equations (6) and (7a, b) become the differential equation

$$\frac{\partial}{\partial y} \left\{ (1-y^2) \frac{\partial v(y, \tau)}{\partial y} \right\} = \frac{\partial^2 v(y, \tau)}{\partial \tau^2} \quad (9)$$

and the conditions

$$\tau=0: v(y, 0)=0, \partial v/\partial \tau|_{\tau=0}=0, 0 \leq y \leq \alpha l < 1, \quad (10a)$$

$$y=0: v(0, \tau)=0; \quad y=\alpha l: v(\alpha l, \tau)=B \sin \beta \tau, \tau > 0, \quad (10b)$$

where

$$\alpha = \left( \frac{\frac{1}{2}\rho A\omega^2}{T_0 + \left\{ \frac{1/6}{1/2} \right\} \rho A\omega^2 l^2} \right)^{\frac{1}{2}}, \quad \beta = \frac{\nu\sqrt{2}}{\omega}. \quad (10c)$$

$$\text{Define } * \bar{v}(y, \lambda) = \int_0^\infty e^{-\lambda \tau} v(y, \tau) d\tau, \quad \text{Re}(\lambda) > 0.$$

---

\* This definition of the Laplace transform of the function  $v(y, \tau)$  in terms of the complex parameter  $\lambda$  assumes that the real part of  $\lambda$  is sufficiently large to render the integral absolutely convergent and that the function  $v(y, \lambda)$  can be so rendered.

Applying the transformation to (9) and (10 *a, b*), we get the subsidiary Legendre equation

$$\frac{d}{dy} \left\{ (1-y^2) \frac{d\bar{v}(y, \lambda)}{dy} \right\} - \lambda^2 \bar{v}(y, \lambda) = 0 \quad \dots \quad (11)$$

subject to the conditions at the ends

$$\left. \begin{aligned} y=0 : \bar{v}(0, \lambda) &= 0, \\ y=\alpha l : \bar{v}(\alpha l, \lambda) &= B\beta/(\lambda^2 + \beta^2) = B \int_0^\infty e^{-\lambda\tau} \sin \beta\tau \cdot d\tau. \end{aligned} \right\} \quad \dots \quad (12)$$

Assume a solution for (11) as

$$\bar{v}(y, \lambda) = \sum_{n=0}^{\infty} a_n(\lambda) \cdot y^n \quad \dots \quad (13)$$

We find the recurrence relation to be

$$a_{n+2} = \frac{\lambda^2 + n(n+1)}{(n+1)(n+2)} a_n.$$

Because of the condition  $\bar{v}(0, \lambda) = 0$ , only odd integers for  $n$  need be considered and the solution is given by the transcendental Legendre function

$$\bar{v}(y, \lambda) = a_1 \left\{ y + \frac{\lambda^2 + 1 \cdot 2}{3!} y^3 + \frac{(\lambda^2 + 1 \cdot 2)(\lambda^2 + 3 \cdot 4)}{5!} y^5 + \dots \right\}.$$

The series

$$V(y, \lambda^2) = y + \frac{\lambda^2 + 1 \cdot 2}{3!} y^3 + \frac{(\lambda^2 + 1 \cdot 2)(\lambda^2 + 3 \cdot 4)}{5!} y^5 + \dots \quad \dots \quad (A)$$

is absolutely convergent for all values of  $\lambda$  when  $|y| < 1$ . By (5) and (8) we are assured the  $V(y, \lambda^2)$  is acceptable, i. e.,  $y < 1$ .

By (12) we have

$$\bar{v}(\alpha l, \lambda) = a_1 V(\alpha l, \lambda^2) = B\beta/(\lambda^2 + \beta^2),$$

so that the solution for the transformed equation (11) is

$$\bar{v}(y, \lambda) = B\beta \frac{V(y, \lambda^2)}{(\lambda^2 + \beta^2)V(\alpha l, \lambda^2)} \quad \dots \quad (14)$$

The Inversion Theorem\* for the Laplace transform gives the solution as the sum of the residues of the contour integral

$$v(y, \tau) = \frac{B\beta}{2\pi i} \oint \frac{e^{\lambda\tau} \cdot V(y, \lambda^2)}{(\lambda^2 + \beta^2) \cdot V(\alpha l, \lambda^2)} d\lambda, \quad \dots \quad (15)$$

---

\* H. S. Carslaw and J. C. Jaeger, 'Operational Methods in Applied Mathematics,' ch. v. (Oxford University Press, 1941).

in which the contour indicated is an infinitely large circle about the origin. We can write

$$\begin{aligned} V(y, \lambda^2) = & \left\{ y + \frac{1}{3}y^3 + \frac{1}{5}y^5 + \dots \right\} \\ & + \lambda^2 \left\{ \frac{1}{3!}y^3 + \frac{1 \cdot 2 + 3 \cdot 4}{5!}y^5 + \dots \right\} \\ & + \lambda^4 \left\{ \frac{1}{5!}y^5 + \frac{1 \cdot 2 + 3 \cdot 4 + 5 \cdot 6}{7!}y^7 + \dots \right\} + \dots \\ = & A_0(y) + \lambda^2 A_2(y) + \lambda^4 A_4(y) + \dots \quad (B) \end{aligned}$$

Let the zeros\* of  $V(\alpha l, \lambda^2)$  be at

$$\lambda^2 = -\lambda_n^2, \quad n=1, 2, 3, \dots$$

Hence the poles, all simple, of the integrand in (15) are

$$\lambda = \pm i\beta, \quad \pm i\lambda_n, \quad n=1, 2, 3, \dots$$

#### Calculation of the Residues.

$$\boxed{\lambda = \pm i\beta}$$

We have as the residues

$$B\beta \cdot \frac{e^{\pm i\beta\tau} \cdot V(y, -\beta^2)}{\pm 2i\beta V(\alpha l, -\beta^2)},$$

and the sum of the residues at  $\pm i\beta$  is simply

$$B \frac{V(y, -\beta^2)}{V(\alpha l, -\beta^2)} \sin \beta\tau.$$

$$\boxed{\lambda = \pm i\lambda_n}$$

The residues, expressed differentially, are

$$B\beta \left[ \frac{e^{\pm i\lambda_n\tau} \cdot V(y, -\lambda_n^2)}{(\beta^2 - \lambda_n^2) \left( \frac{dV(\alpha l, \lambda^2)}{d\lambda} \right) \Big|_{\lambda = \pm i\lambda_n}} \right].$$

Let

$$\begin{aligned} dV(\alpha l, \lambda^2)/d\lambda \Big|_{\lambda = \pm i\lambda_n} &= \pm iW(\alpha l, \lambda_n), \\ W(\alpha l, \lambda_n) &= 2\lambda_n A_2(\alpha l) - 4\lambda_n^3 A_4(\alpha l) + 6\lambda_n^5 A_6(\alpha l) - \dots \quad (C) \end{aligned}$$

\* It is clear  $V(\alpha l, \lambda^2)$  can have no real zeros. Neither can it have any complex zeros because, if  $p \pm ig$  were zeros for  $V(\alpha l, \lambda^2)$ , then the residues contributed at  $p \pm ig$  would introduce an exponential factor  $e^{g\tau}$  in time which would mean the solution went to infinite values *exponentially* in time, contrary to fact. Moreover,  $-p \pm ig$  cannot be zeros because so then would  $p \pm ig$ . Hence  $V(\alpha l, \lambda^2)$  can have only purely imaginary zeros. The transverse displacement, however, can go to large values with time for the special case of resonance without damping, and this is discussed in a special section.

The use of the *transcendental Legendre functions* of real or imaginary order is indeed rather novel in a practical problem and, as a consequence, the study of their zeros has been neglected by mathematicians until recently. See, for example, Ince, 'Ordinary Differential Equations', ch. xxi.



Then, since

$$\left. \frac{dV(\alpha l, \lambda^2)}{d\lambda} \right|_{\lambda=i\lambda_n} = - \left. \frac{dV(\alpha l, \lambda^2)}{d\lambda} \right|_{\lambda=-i\lambda_n},$$

the sum of the residues at  $\pm i\lambda_n$  is found to be

$$2B\beta \frac{V(y, -\lambda_n^2)}{(\beta^2 - \lambda_n^2) \cdot W(\alpha l, \lambda_n)} \sin \lambda_n \tau.$$

Adding all the residues, we obtain the solution

$$v(y, \tau) = B \frac{V(y, -\beta^2)}{V(\alpha l, -\beta^2)} \sin \beta \tau + 2B\beta \sum_{n=1}^{\infty} \frac{V(y, -\lambda_n^2)}{(\beta^2 - \lambda_n^2) \cdot W(\alpha l, \lambda_n)} \sin \lambda_n \tau \quad (16)$$

for the transverse displacement of the wire, where the functions  $V$  and  $W$  are defined in (A), (B) and (C) and  $\alpha$ ,  $\beta$ , and  $\tau$  by (8) and (10 c).

It is not difficult to show that the set of eigen-functions  $\{V(y, -\lambda_n^2)\}$  which are obtained for any particular case of the problem considered cannot be an orthogonal set. Let  $V(y, -\lambda_1^2)$  and  $V(y, -\lambda_2^2)$  be any two distinct non-trivial solutions corresponding to the values  $-\lambda_1^2$  and  $-\lambda_2^2$  of the parameter for the equation (11). We obtain readily that

$$(\lambda_1^2 - \lambda_2^2) \int_0^y V_1 \cdot V_2 dy = \left[ (1-y^2) \{V_1 V_2' - V_1' V_2\} \right]_0^y \\ = (1-y^2) \{V(y, -\lambda_1^2) \cdot V'(y, -\lambda_2^2) - V'(y, -\lambda_1^2) \cdot V(y, -\lambda_2^2)\}$$

(since the contribution at the lower limit is zero) where we have written  $V_1 = V(y, -\lambda_1^2)$  and  $V_2 = V(y, -\lambda_2^2)$  and the prime indicates differentiation with respect to  $y$ . The requirement that  $V_1$  and  $V_2$  be orthogonal on or in the range  $0 \leq y \leq \alpha l < 1$  would lead to the vanishing of the Wronskian,  $V_1 V_2' - V_1' V_2$ , since  $1-y^2 > 0$ . But the Wronskian cannot vanish if the functions are linearly independent in  $y$ . Hence they cannot be orthogonal for any value of  $y$ .

As a concluding remark to this section, we may leave the reader to consider the possibility and significance of at least one eigenfunction being a Legendre polynomial of odd order ( $\lambda^2 + n(n+1) = 0$ ).

### Resonance.

Now suppose that for some  $\lambda_k$ ,  $\lambda_k^2 = \beta^2$ . This requires a re-examination of the formal solution (15) because the solution (16) fails for such a case of resonance. Since we assume that  $V(\alpha l, -\beta^2) = V(\alpha l, -\lambda_k^2) = 0$ , we can write

$$V(\alpha l, \lambda^2) = (\lambda^2 + \beta^2)(b_0 + b_2 \lambda^2 + b_4 \lambda^4 + \dots) \\ = (\lambda^2 + \beta^2) \cdot \Psi(\alpha l, \lambda^2) \\ = b_0 \beta^2 + (b_0 + b_2 \beta^2) \lambda^2 + (b_2 + b_4 \beta^2) \lambda^4 + \dots$$

This gives

$$b_0 = A_0/\beta^2,$$

$$b_2 = A_2/\beta^2 - A_0/\beta^4,$$

$$b_4 = A_4/\beta^2 - A_2/\beta^4 + A_0/\beta^6, \text{ etc.},$$

and

$$\Psi(\alpha l, \lambda^2) = b_0 + b_2 \lambda^2 + b_4 \lambda^4 \oplus \dots \quad (D)$$

The formal solution (15) then becomes

$$v(y, \tau) = \frac{B\beta}{2\pi i} \oint \frac{e^{\lambda\tau} \cdot V(y, \lambda^2)}{(\lambda^2 + \beta^2)^2 \cdot \psi(\alpha l, \lambda^2)} d\lambda \quad (17)$$

and the problem of resonance has been reduced to the evaluation of residues at double poles as well as simple poles. We have, therefore, residues at  $\lambda = \pm i\beta = \pm i\lambda_k$  (double poles) and simple poles at  $\lambda = \pm i\lambda_n$ ,  $n \neq k$ ,  $n = 1, 2, 3, \dots, k-1, k+1, \dots$ . It is only necessary to determine the residues at the double poles. These are

$$\begin{aligned} & B\beta \frac{d}{d\lambda} \left[ \frac{e^{\lambda\tau} \cdot V(y, \lambda^2)}{(\lambda \pm i\beta)^2 \cdot \psi(\alpha l, \lambda^2)} \right]_{\lambda = \pm i\beta = \pm i\lambda_k} \\ &= B\beta e^{\pm i\lambda_k \tau} \left[ \frac{-\tau \mp iW(y, \lambda_k)}{4\beta^2 \cdot \psi(\alpha l, -\lambda_k^2)} \mp \frac{iV(y, -\lambda_k^2)}{4\beta^3 \psi(\alpha l, -\lambda_k^2)} \pm \frac{iV(y, -\lambda_k^2) \cdot U(\alpha l, \lambda_k)}{4\beta^2 \psi^2(\alpha l, -\lambda_k^2)} \right] \end{aligned}$$

after some differentiation, where

$$\pm iU(\alpha l, \lambda_k) = \frac{d\psi(\alpha l, \lambda^2)}{d\lambda} \Big|_{\lambda = \pm i\lambda_k} \quad (E)$$

The sum of the residues at  $\lambda = \pm i\beta = \pm i\lambda_k$  is found to be

$$\begin{aligned} & \frac{B}{2\lambda_k} \left[ \left( \frac{W(y, \lambda_k)}{\psi(\alpha l, -\lambda_k^2)} + \frac{V(y, -\lambda_k^2)}{\lambda_k \psi(\alpha l, -\lambda_k^2)} \right. \right. \\ & \quad \left. \left. - \frac{V(y, -\lambda_k^2) \cdot U(\alpha l, \lambda_k)}{\psi^2(\alpha l, -\lambda_k^2)} \right) \sin \lambda_k \tau - \tau \cdot \cos \lambda_k \tau \right]. \end{aligned}$$

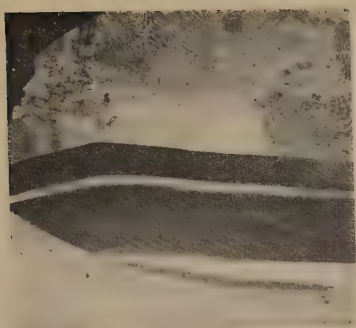
The solution for the case of resonance is then

$$\begin{aligned} v(y, \tau) &= \frac{B}{2\lambda_k} \left[ \left( \frac{W(y, \lambda_k)}{\psi(\alpha l, -\lambda_k^2)} + \frac{V(y, -\lambda_k^2)}{\lambda_k \cdot \psi(\alpha l, -\lambda_k^2)} \right. \right. \\ & \quad \left. \left. - \frac{\psi^2(\alpha l, -\lambda_k^2)}{V(y, -\lambda_k^2) \cdot U(\alpha l, \lambda_k)} \right) \sin \lambda_k \tau - \tau \cdot \cos \lambda_k \tau \right] \\ & \quad + 2B\lambda_k \sum_{n=1}^{\infty} (n \neq k) \frac{V(y, -\lambda_n^2)}{(\lambda_k^2 - \lambda_n^2) \cdot W(\alpha l, \lambda_n)} \cdot \sin \lambda_n \tau, \quad (18) \end{aligned}$$

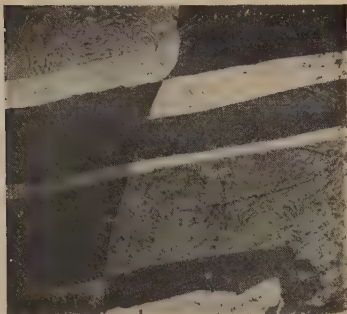
and we see how the transverse displacement goes to large values with time in resonance.

---

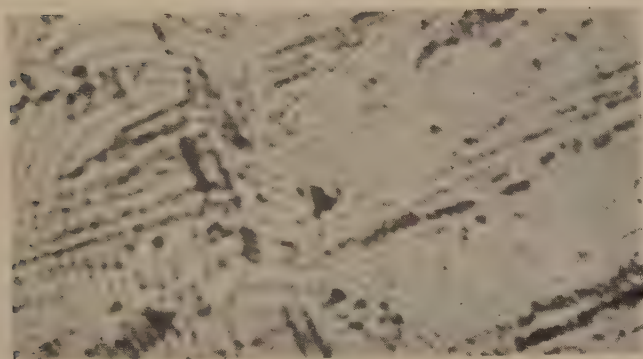
[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



(a)



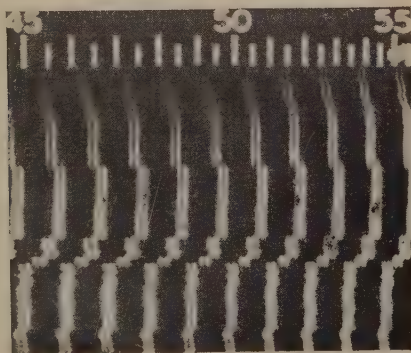
(b)



(c)



(d)



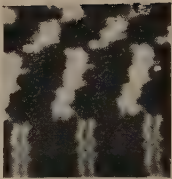
(e)



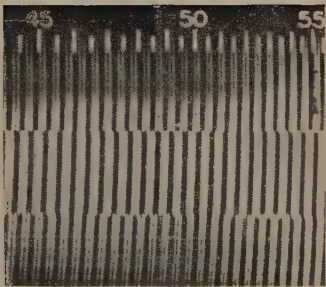
(a)



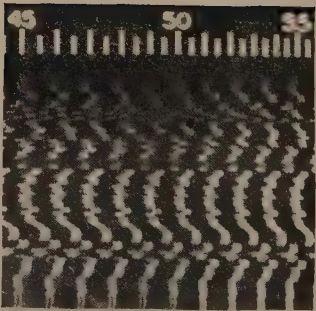
(c)



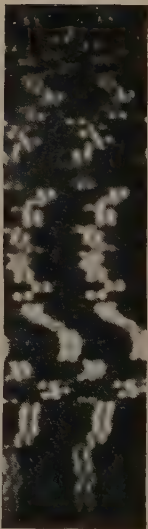
(f)



(b)

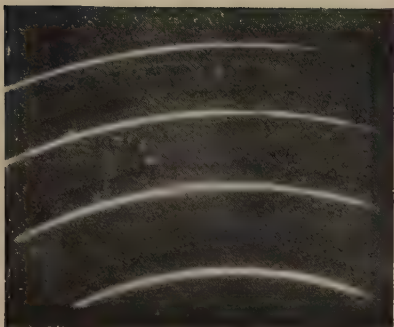


(d)



(e)

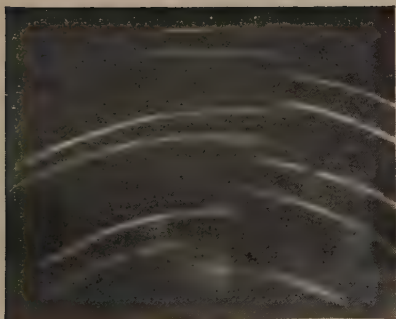




(a)



(b)



(c)



(d)



FIG. 9.

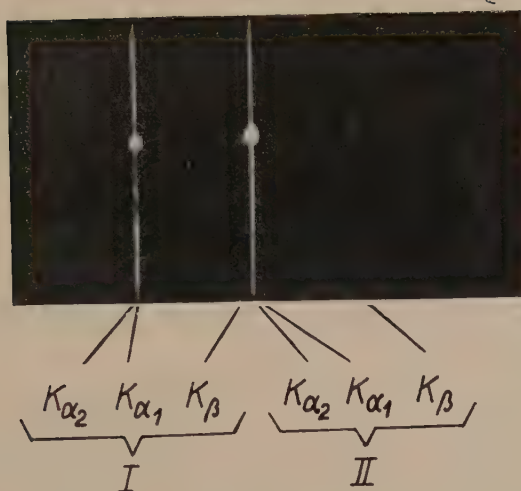
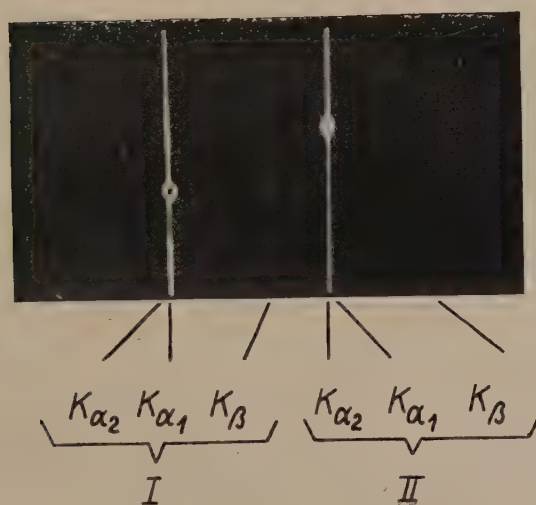


FIG. 10.







### XLIII. *Theory of the High Pressure Mercury Vapour Discharge.*

By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E.

(Communication from the Staff of the Research Laboratories of  
The General Electric Company Limited, Wembley, England.) \*

[Received December 10, 1945.]

#### 1. *Introduction.*

The theory of the positive column of the high-pressure mercury vapour discharge was developed by Elenbaas <sup>(1), (2), (3), (4), (5)</sup>. By a semi-experimental method <sup>(1)</sup> he found the temperature of the column for a particular discharge, and later, using this temperature, calculated other parameters of the discharge. By this process and later <sup>(3)</sup> by studying the properties of discharges obeying "similarity laws" he avoided an attempt to solve the energy balance equation. His results, therefore, although of very great value as providing almost the only theoretical basis we have for the high pressure discharge, are of somewhat restricted interest. Only by means of the solution of the energy balance equation can relations between parameters be studied for a wide range of pressures and operating characteristics; and only by this means can the validity of the basic assumptions made be estimated.

Now the solution of an energy balance equation is always difficult—partly because of its complicated nature and partly because the values or constants needed are rarely known—and, so far as the writer is aware, no attempt has been made to solve the equation for the high pressure mercury vapour discharge. An attempt is made here, however, to find such a solution.

#### 2. *The Basic Assumptions and The Energy Balance Equation.*

It is best to study the assumptions which we make as the energy balance equation is set up. Consider a positive column lamp operating vertically. It is taken to be so long that any end effects both as regards loss of heat and also any variations in vapour density due to convection are negligible. All conditions are supposed uniform along the vertical direction of the column, so that we can consider one centimetre length of column with no end effects as being representative of the whole discharge. The radius of the discharge tube is  $r_0$  cm.; the discharge has a potential drop  $E$  per cm. of arc at current  $I$ ; it is assumed that  $E$  does not vary over the cross-section of the discharge; this is likely to be very nearly true. At a distance  $r$  from the axis of the discharge the current density is  $i(r)$ .

---

\* Communicated by Sir Clifford Paterson, F.R.S.



where  $(r)$  is the electron density,  $e$  the electron charge and  $u$  the mobility given by

$$u = \frac{\alpha e \lambda}{\sqrt{m_e k T}}, \quad . . . . . (2.7)$$

$\alpha$  being a constant of approximate value 0.5; its value depends on the method of averaging used in deriving the formula; we do not need to know its actual value for reasons which will appear.  $\lambda$  is the electron mean free path,  $m_e$  the electron mass,  $k$  Boltzmann's constant. If  $\lambda_0$  is the electron mean free path reduced to N.T.P. we find

$$u = \frac{1.22 \times 10^{21} \cdot \alpha \lambda_0}{n(r) \sqrt{T}}. \quad . . . . . (2.8)$$

All the quantities are measured in E.M.C.G.S. units. The numerical coefficient in (2.8) involves only constants which are known accurately. We do not need to use a value for  $\lambda_0$ , which is fortunate since its value is known only roughly.

From (2.6) and (2.8) it follows at once that

$$i(r) = 19.4 \cdot \alpha \lambda_0 \frac{x(r) \cdot E}{\sqrt{T}}, \quad . . . . . (2.9)$$

where  $x(r)$  is the degree of ionization.  $x(r)$  may be expressed in terms of  $T(r)$  and  $P$ , the pressure, by means of Saha's equation, the form of which, applicable to the case considered here is

$$x = 10^{-\left(\frac{2.63 \times 10^4}{T} + 3.25\right)} P^{0.5} T^{1.25}, \quad . . . . . (2.10)$$

where  $P$  is measured in atmospheres.

Combining (2.10) and (2.9) we have

$$i = 1.09 \times 10^{-2} \cdot \alpha \lambda_0 \cdot E P^{-\frac{1}{2}} \cdot T^{\frac{3}{2}} 10^{-\frac{2.63 \times 10^4}{T}}. \quad . . . . . (2.11)$$

The function  $a(n, T)$  is more difficult to deal with. Elenbaas, in his earlier papers, treats  $a(n, T)$  as constant. It is, in fact, so far from being constant that it probably varies by a factor of several hundred throughout the range of values we are concerned with.

The energy lost by line radiation is clearly proportional to  $nf(T)$ , where the unknown function  $f(T)$  involves the values of the energy levels and their quantum weights. Where any appreciable amount of the radiation is continuous it may be no longer given by an expression of this form, but we make the assumption (probably not strictly true in practice) that the effect of continuous radiation is negligible. So far as line radiation alone is concerned, if only one excited level of voltage  $V_M$  were involved, we could write

$$a_1(n, T) = A_1 n \exp \left\{ -\frac{11600 V_M}{T} \right\}, \quad . . . . . (2.12)$$

where  $A_1$  is a constant involving the quantum weight of the level and the transition probability. Elenbaas <sup>(4)</sup>, in one of his later papers, made the





any two of the quantities  $M$ ,  $I$ ,  $E$ ,  $P$  (within the range of reasonable values) may be chosen at will, and will then determine the other two. In this determination  $r_0$  and  $T_w$  will, of course, appear as parameters. We should therefore be able to deduce from (2.14) the relations

$$M=M(E, P, r_0, T_w), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

$$I=I(E, P, r_0, T_w). \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.2)$$

It will be shown that this is, in fact, the case. The important point, however, is that  $E$  and  $P$  definitely fix the conditions in the discharge. Equation (2.14) for these conditions when solved expresses  $T$  as a function of  $r$  ( $r_0$  and  $T_w$  appearing as constants in the integration), from which function all the other properties of the discharge may be calculated.

The major difficulty in the solution of (2.14) is that we do not want to make use of calculated values of the constants  $K_1$  and  $A_3$  but wish to determine their best values by reference to certain measurable properties of the high pressure mercury vapour discharge itself. We therefore have to seek for the solution of this equation with  $K_1$  and  $A_3$  as parameters. We find it necessary, of course, to use some approximate method. The method used will appear in detail below, but in general terms it is as follows: consideration of the function on the right of (2.14) (which is proportional to the thermal energy developed in the vapour) shows that it increases rapidly with temperature over the region in which we are interested. This is, of course, the cause of the familiar "constriction" of the arc. We therefore determine a radius  $r_1$  corresponding to a temperature  $T_1$  at which the value of the function

$$\mu\left(T, \frac{A_3 P^{3/2}}{E^2}\right) = \lambda(T) - \frac{A_3 P^{3/2}}{E^2} \phi(T) \quad . \quad . \quad . \quad . \quad . \quad (3.3)$$

first becomes significant. We then, to obtain an approximate form for the temperature distribution, assume that within the radius  $r_1$  the entire energy is evolved and that it is uniformly distributed, and that outside  $r_1$  no evolution of energy takes place. This is the only serious approximation made in the solution of the energy balance equation as distinct from the assumptions made in deriving it. It can be a bad approximation if the thermal energy as a function of  $r$  is peaky near  $r=0$  and, in considering the results obtained, due account will have to be taken of the fact that this device has been used.

When the thermal energy evolved has a strong maximum along the axis of the discharge, the general effect of the error so introduced will be to give a temperature distribution over the discharge flatter than that actually occurring. In other cases the approximation is not at all a bad one. In any particular case it is always possible if required to obtain a second approximation by substituting the first approximation in the right-hand side of equation (2.14) and solving again. For our purpose here, which is to obtain from energy balance considerations a general representation of the behaviour of the discharge the first approximation will suffice.

We may now write (2.14) as

$$\left. \begin{aligned} \frac{1}{r} \frac{d}{dr} \left\{ r \frac{dT^2}{dr} \right\} &= -K_1 E^2 P^{-\frac{1}{2}} \mu_0 \quad (r < r_1) \\ &= 0 \quad (r > r_1) \end{aligned} \right\}, \quad \dots \quad (3.4)$$

where  $\mu_0$  is independent of  $r$ , and where, in order that the energy evolved within the circle of radius  $r_1$  should be equal to the actual energy evolved, we must have

$$\pi r_1^2 K_1 E^2 P^{-\frac{1}{2}} \mu_0 = \int_0^{r_0} 2\pi r K_1 E^2 P^{-\frac{1}{2}} \mu \left( T, \frac{A_3 P^{3/2}}{E^2} \right) dr \quad \dots \quad (3.5)$$

or 
$$\mu_0 = \frac{1}{r_1^2} \int_0^{r_0} 2r \mu(T) dr. \quad \dots \quad (3.6)$$

We may now write down the solution of (3.4) as

$$\left. \begin{aligned} T^2 &= T_0^2 - \frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0}{4} r^2 \quad (r < r_1) \\ T^2 &= T_w^2 + B \ln \frac{r_0}{r} \quad (r > r_1) \end{aligned} \right\}, \quad \dots \quad (3.7)$$

where  $T_w$  and  $T_0$  are the temperatures at the wall and the axis of the discharge respectively, and  $B$  is a constant. To find  $B$  we have the boundary conditions that  $T$  and  $dT/dr$  are continuous at  $r=r_1$ , which gives us

$$B = \frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{2} \quad \dots \quad (3.8)$$

and, in addition, the relations

$$T_1^2 = T_0^2 - \frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0}{4} r_1^2, \quad \dots \quad (3.9)$$

$$T_1^2 = T_w^2 + \frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{2} \ln \frac{r_0}{r_1}. \quad \dots \quad (3.10)$$

Now when the conditions of the discharge are specified—which, as we have seen, is equivalent to fixing  $E$  and  $P$ —and the constants  $K_1$  and  $A_3$  are known,  $\mu_0$  and  $T_1$  are fixed in terms of  $E$  and  $P$  although the forms of these functions are not yet determined. The equations (3.9) and 3.10) are then sufficient to determine  $r_1$  and  $T_0$ , and these values substituted in equations (3.7) give the required solution. We now proceed to find the forms of  $T_1$  and  $\mu_0$ .

#### 4. The Functions $\mu_0$ and $T_1$ .

In order to consider the function  $\mu$  given by (3.3), we need an approximation to the actual value of  $A_3$ . It doesn't really matter what value of  $A_3$  we assume; by a process of trial and error we should find whether we are right or wrong, but the more nearly we can make a correct guess at  $A_3$  the shorter will be the work involved. It would be possible

to make an estimate of  $A_3$  by calculation. We have available, however, two pieces of information which allow us easily to form an approximation.

The first is that Elenbaas <sup>(1)</sup> found that in a particular type of lamp with  $E=8$  volts and  $P=1$  atmosphere the maximum temperature was of the order of  $6500^\circ\text{K}$ . The second is that in these circumstances it is known that something like 50 per cent. of the energy is radiated from the discharge <sup>(6)</sup>. We therefore take, for  $P=1$ ,  $E=8 \times 10^8$   $T=6000$ ,

$$\frac{A_3 P^{3/2}}{E^2} \frac{\phi(T)}{\lambda(T)} = 0.5, \quad \dots \dots \dots (4.1)$$

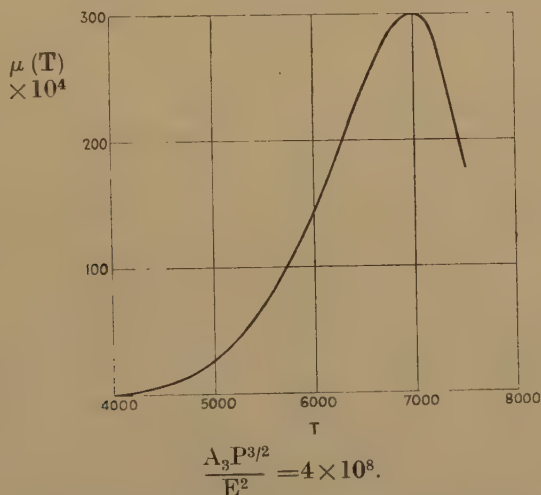
which gives approximately

$$A_3 = 2.5 \times 10^{26}. \quad \dots \dots \dots (4.2)$$

Now, for the range of operating conditions in which we are mainly interested,  $P$  varies from 1 to 10 atmospheres while  $E$  varies from 5 to 100 volts, and in general the larger values of  $E$  are associated with the larger values of  $P$ , therefore the range of  $A_3 P^{3/2}/E^2$  will be from some  $4 \times 10^8$  to  $10^9$ .

The function  $\mu(T)$  for  $A_3 P^{3/2}/E^2 = 4 \times 10^8$  is shown in fig. 1. The general shape of the function is the same for other values of the parameter  $A_3 P^{3/2}/E^2$ , but the value of  $T$ , at which the peak ( $T_P$ ) occurs as well as the

Fig. 1.



value of  $\mu(T)$  at this peak ( $\mu_P$ ), depends very much on this parameter. Since, as is shown below, it was necessary to plot a number of these curves, and as they are not all given here, the values of  $T_P$  and  $\mu_P$  are given for four values of  $A_3 P^{3/2}/E^2$  in Table I.

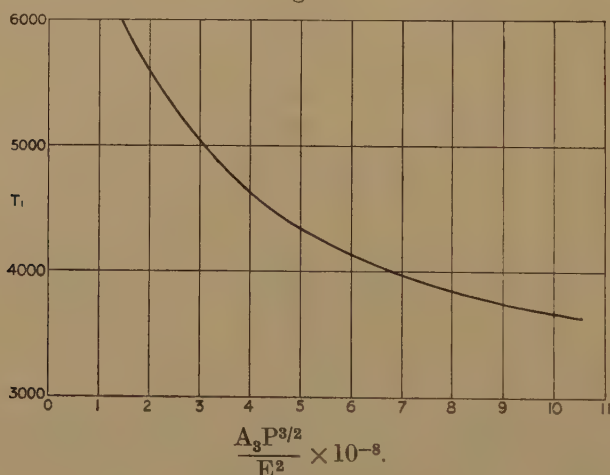
Having plotted out the curves corresponding to the four values of  $A_3 P^{3/2}/E^2$  in Table I., it is possible to decide on  $T_1$  as a function of  $A_3 P^{3/2}/E^2$ . For this purpose we may choose  $T_1$  to some extent to suit our convenience, subject to the condition mentioned in Section 3, that nearly all the energy must be dissipated inside the radius corresponding to the value of  $T_1$ .

TABLE I.

$\frac{A_3 P^{3/2}}{E^2}$	$T_p$	$\mu_p$
$2 \times 10^8$	9800	$4.8 \times 10^{-1}$
$4 \times 10^8$	7000	$3.0 \times 10^{-2}$
$6 \times 10^8$	5900	$6.9 \times 10^{-3}$
$10 \times 10^8$	5300	$1.8 \times 10^{-3}$

chosen, and that  $T_1$  must be a continuous function of  $A_3 P^{3/2}/E^2$  in order to allow interpolation.  $T_1$  must not be too small, however, otherwise the approximation we have made in deriving equations (3.7) becomes more serious. Examination of the curves of  $\mu(T)$  lead us to choose the values of  $T_1$  given in fig. 2. For example, for the value of  $A_3 P^{3/2}/E^2 = 4 \times 10^8$ , as shown in fig. 1,  $T_1 = 4600$ , the value at which  $\mu(T)$  begins to increase rapidly with increasing  $T$ .

Fig. 2.



We must now obtain  $\mu_0$  as a function of  $E$  and  $P$ . For this purpose we go back to equation (3.6). Since  $\mu(T)$  is negligible for values of  $r > r_1$  (where, we remember,  $r_1$  is the value of  $r$  corresponding to  $T_1$ ), we may write

$$\left. \begin{aligned} \mu_0 &= \frac{1}{r_1^2} \int_0^{r_1} 2r\mu(T)dr \\ &= \frac{1}{r_1^2} \int_0^{r_1^2} \mu(T)d(r^2) \end{aligned} \right\}, \dots \dots \dots (4.3)$$

and using (3.7) we get

$$\mu_0 = \frac{4}{K_1 E^2 P^{-1} \mu_0 r_1^2} \int_{T_1^2}^{T_0^2} \mu(T)d(T^2). \dots \dots \dots (4.4)$$

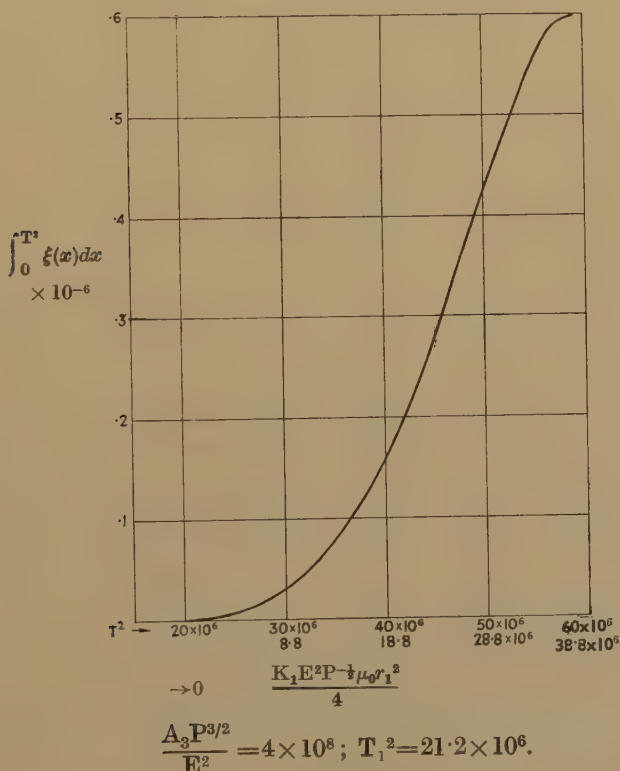


If we write  $\mu(T) = \xi(T^2)$  and use (3.9), (4.4) becomes

$$\mu_0 = \frac{4}{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2} \int_{T_1^2}^{T_1^2 + \frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4}} \xi(T^2) d(T^2). \quad (4.5)$$

Since  $T_1$  is known as a function of  $A_3 P^{3/2}/E^2$  we may obtain the value of the integral as a function of  $K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2/4$  for any given value of  $A_3 P^{3/2}/E^2$ . Taking the case  $A_3 P^{3/2}/E^2 = 4 \times 10^8$  as example, the curve of fig. 1 is replotted with  $T^2$  as abscissæ, numerical integration then gives us at once  $\int_0^{T_1^2} \xi(x) dx$ , which is shown in fig. 3; it will be noticed that the curve becomes horizontal above  $T^2 = 60 \times 10^6$ . This feature is common and is due to the fact that  $\mu(T)$  falls to zero. The value of  $T^2$  at which this happens is, of course, a function of  $A_3 P^{3/2}/E^2$ . If, in fig. 3, we subtract  $T_1^2$  (which in this case is  $21.2 \times 10^6$ ) from the abscissæ the resulting curve is the integral on the right of (4.5) as a function of  $K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2/4$ .

Fig. 3.



If now, according to equation (4.5), we divide each value of the integral in fig. 3 by the corresponding value of  $K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2/4$ , we obtain for that value of the abscissæ the corresponding value of  $\mu_0$ . This is shown plotted in fig. 4 (b). We might go further and from the corresponding

values of ordinate and abscissa in fig. 4 (b) obtain  $\mu_0$  as a function of  $K_1 E^2 P^{-1} r_1^2/4$ , but for reasons which will appear it is convenient to leave the relation in the form given in fig. 4 (b). This same process was followed for the four values of  $A_3 P^{3/2}/E^2$  given in Table I., and the resulting curves are shown in figs. 4 (a), 4 (b), 4 (c), 4 (d). These curves are reproduced since they represent important functions in the theory.

Fig. 4 (a).

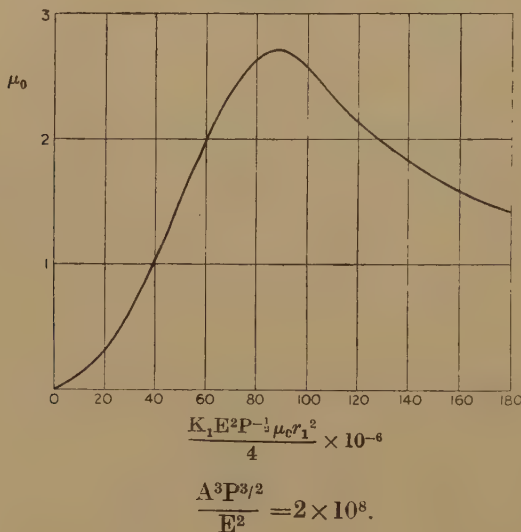


Fig. 4 (b).

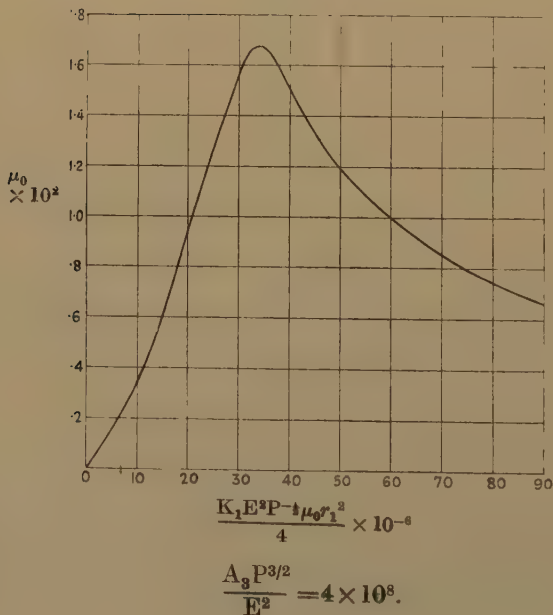


Fig. 4 (c).

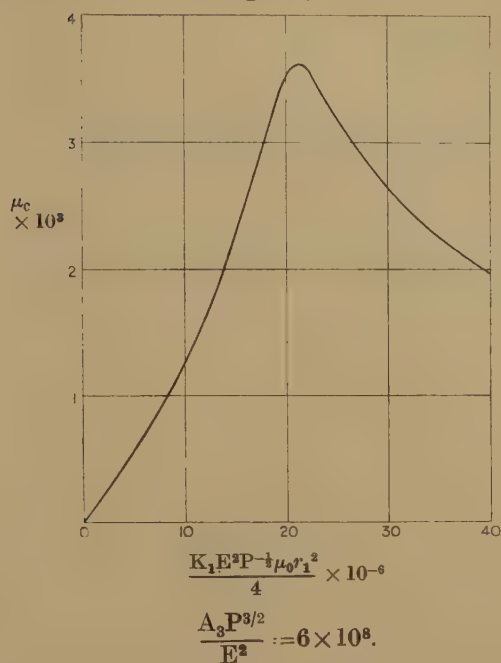
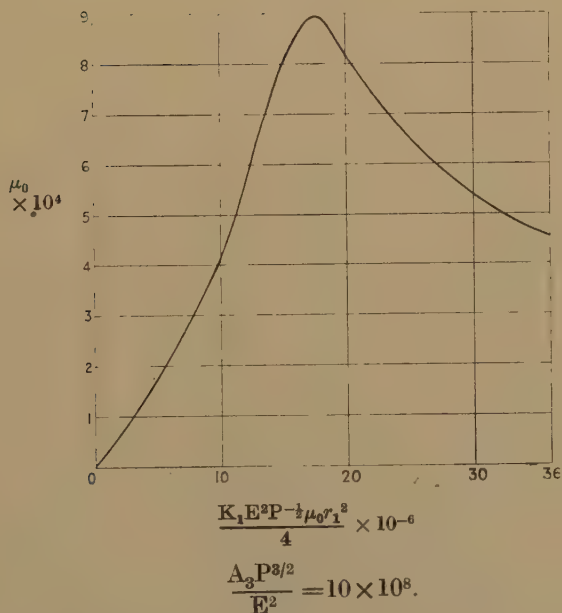


Fig. 4 (d).



We now have formally, in figs. 2 and 4,  $T_1$  as a function of  $A_3 P^{3/2}/E^2$  and  $\mu_0$  as a function of  $K_1 E^2 P^{-\frac{1}{2}} r_1^2/4$ . If  $A_3$ ,  $K_1$  were known and  $P$ ,  $E$  were given, substitution of these two functions in (3.9) and (3.10) would be





for small values of  $t$  the contribution to the integral is small. The error introduced by replacing for values of  $t < r_1/r_0$  the integrand of the second integral by that of the first will, therefore, be inappreciable and we may write

$$M = \frac{15 \cdot 3 r_0^2 P}{T_w} \int_0^1 \frac{t dt}{\sqrt{1 - \frac{K_1 E^2 P^{-1} \mu_0 r_1^2}{4 T_w^2} \ln t}} \quad (5.5)$$

This device is not essential. It would be possible to deal with each integral separately in the same way as the integral in (5.5) is treated, but the extra complication is not justified by the small increase in accuracy achieved.

Fig. 6 (a).

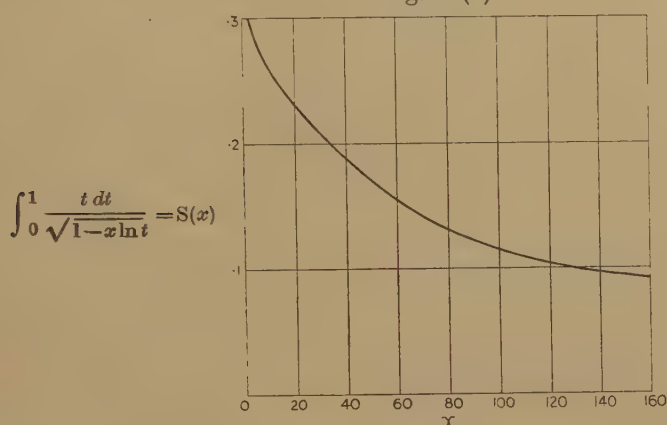
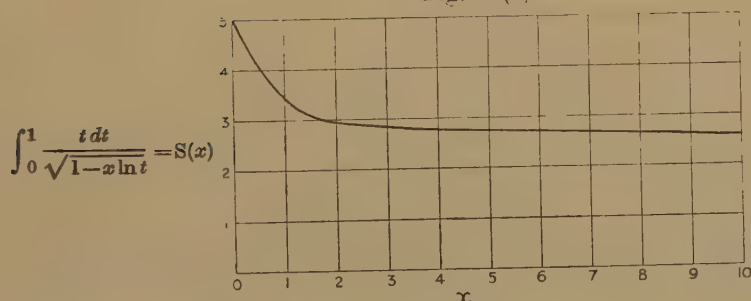


Fig. 6 (b).



The curves in fig. 5 were integrated numerically and the value of the integral  $\int_0^1 \frac{t dt}{\sqrt{1 - x \ln t}}$  obtained as a function of  $x$ .

If we write this function as  $S(x)$  we have

$$M = \frac{15 \cdot 3 r_0^2 P}{T_w} S \left( \frac{K_1 E^2 P^{-1} \mu_0 r_1^2}{4 T_w^2} \right) \quad (5.6)$$

$S(x)$  is plotted in fig. 6. We have therefore introduced one more variable  $M$  and one other equation, but no other unknown constant.

6. The Equation for I.

From (2.11) and (2.15) we have

$$i=1.09 \times 10^{-2} \alpha \lambda_0 EP^{-\frac{1}{2}} \lambda(T), \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

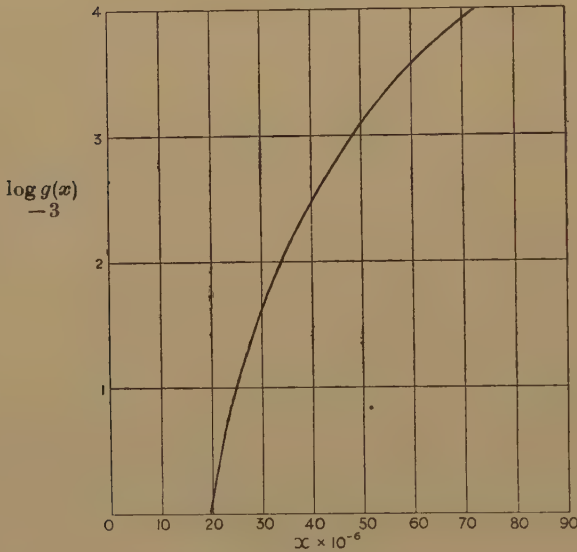
whence

$$I=2\pi \times 1.09 \times 10^{-2} \alpha \lambda_0 EP^{-\frac{1}{2}} \int_0^{r_0} r \lambda(T) dr. \quad . \quad . \quad . \quad . \quad . \quad (6.2)$$

Now  $\lambda(T)$  is negligible for values of  $T$  corresponding to  $r > r_1$ . Accordingly, we may replace the upper limit of the integral in (6.2) by  $r_1$  and use the first equation of (3.7) to transform the variable. We then have

$$I=0.137 \left( \frac{\alpha \lambda_0}{K_1} \right) \frac{1}{E \mu_0} \int_{T_1^2}^{T_0^2} \lambda(T) d(T^2). \quad . \quad . \quad . \quad . \quad . \quad (6.3)$$

Fig. 7.



Again, we may replace, without sensible error, the lower limit  $T_1^2$  by 0, and if we write

$$\eta(T^2)=\lambda(T) \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.4)$$

(6.3) becomes

$$I=0.137 \left( \frac{\alpha \lambda_0}{K_1} \right) \frac{1}{E \mu_0} \int_0^{T_0^2} \eta(y) dy. \quad . \quad . \quad . \quad . \quad . \quad (6.5)$$

The integral in (6.5) was obtained as a function of the upper limit by numerical integration. We write

$$g(x)=\int_0^x \eta(y) dy, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.6)$$

and the function  $\log g(x)$  is shown plotted in fig. 7. Accordingly, we write (6.5) as

$$I = 0.137 \left( \frac{\alpha \lambda_0}{K_1} \right) \frac{1}{E \mu_0} g(T_0^2). \quad (6.7)$$

In obtaining this equation we have introduced what is effectively a new constant, namely  $\alpha \lambda_0 / K_1$ , whose value we wish to find. We are now in a position to attempt to find values for  $A_3$ ,  $K_1$  and  $\alpha \lambda_0 / K_1$ .

### 7. Determination of the Constants.

The parameters, the values of which are easily accessible, are  $M$ ,  $E$  and  $I$ .  $T$  and  $r_1$  cannot be obtained directly. It would be a simple matter to determine values for the constants if we had two accurate sets of  $M$ ,  $E$ ,  $P$ ,  $I$  for two different values of  $P$ . Since the equations do not give an exact representation of the conditions in the discharge, however, it does not follow that values of the constants so obtained will necessarily be the best choice. The best values would be those which gave most nearly agreement with experiment over a wide range of  $P$ ,  $I$  and  $E$ . It is not easy to obtain such optimum values, and all that we shall attempt in this paper is to show that approximate values obtained from experimental data reproduce qualitatively the behaviour of the discharge.

The only time when a set  $M$ ,  $E$ ,  $P$ ,  $I$  can be found without considerable difficulty is when  $P=1$ . This has been done by Elenbaas <sup>(1)</sup> and we have the set of approximate values.

$$\left. \begin{aligned} M &= .003 \text{ grams,} \\ E &= 8 \text{ volts,} \\ r_0 &= 1.0 \text{ cm.,} \\ P &= 1 \text{ atmosphere,} \\ I &= 5.5 \text{ amp.} \end{aligned} \right\} \dots \dots \dots (7.1)$$

We assume from now on that  $T_\omega = 1000^\circ \text{ K}$ . This is approximately true in all practical lamps, and a variation of a few hundred degrees has a small effect on the conditions in the discharge.

We now assume reasonable values for  $A_3$  and find values of  $K_1$  which will allow (5.6) and the function  $\mu_0$  given by figs. 4 (a), 4 (b), 4 (c) and 4 (d) to be satisfied simultaneously; in this way we obtain a relation between  $A_3$  and  $K_1$  which effectively removes one of our unknown constants. Thus we choose a value of  $A_3$  and from (7.1) at once find  $A_3 P^{3/2} / E^2$ . Then from (5.6) and fig. 6 we determine  $\frac{K_1 E^2 P^{-1/2} \mu_0 r_1^2}{4 T_\omega^2}$ . We

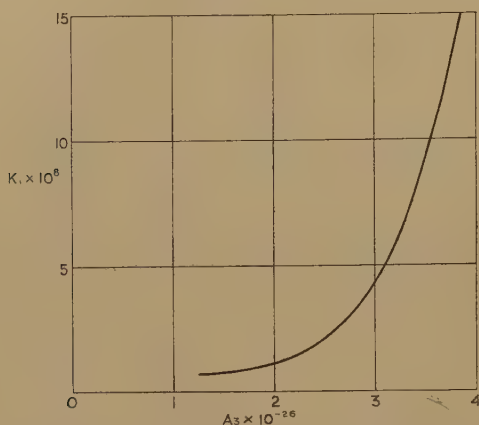
obtain  $\mu_0$  from fig. 4,  $T_1$  from fig. 2 and then  $r_1$  from (3.10). We now have

the value of  $\frac{K_1 E^2 P^{-1/2} \mu_0 r_1^2}{4 T_\omega^2}$  and all the quantities therein except  $K_1$ . We

therefore find the value of  $K_1$  corresponding to the value of  $A_3$  chosen. We can thus find a relation between  $K_1$  and  $A_3$  which must hold if (7.1) is

to be satisfied. This function is shown in fig. 8. To obtain another equation between  $A_3$  and  $K_1$ , and incidentally to check that the relation

Fig. 8.



already found is reasonable, we use the E, I characteristic for the commercial 400 watt lamp. A characteristic of a lamp of this type for which  $r_0 = 1.5$  cm.  $M = 0.012$  gram is shown in fig. 9 (a). To deduce the E, I

Fig. 9 (a).

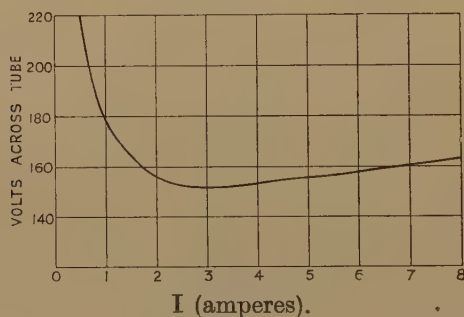
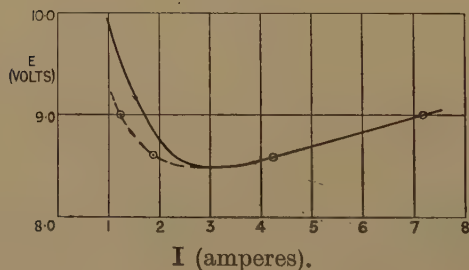


Fig. 9 (b).



curve it is necessary to know the losses at the electrodes. Very little is known about these losses and how they vary with the watts per centimetre



of the discharge column. At the rated value of 400 watts the total electrode losses are likely to be about 40 watts, corresponding to a loss of voltage of about 16 volts. We assume for the moment that this voltage is independent of the tube wattage, and from fig. 9 (a) deduce the E, I relation shown by the full line in fig. 9 (b).

We now use the minimum of this curve to find the correct pair of values of  $A_3$  and  $K_1$ . For this we choose a pair of values of  $A_3$  and  $K_1$  and the value of E at which the minimum occurs, namely 8.5 volts. For example, Table II. shows the calculations for  $A_3=3.4 \times 10^{26}$ ,  $K_1=8.0 \times 10^{-8}$ .

These quantities, of course, satisfy the relation of fig. 8.

TABLE II.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
P	$\frac{A_3 P^{3/2}}{E^2}$	$T_1^2$	$\frac{K_1 E^2 P^{-1/2} \mu_0 r_1^2}{4 T_\omega^2}$	$\mu_0$	$r_1$	$K_1$
1.3	$7.0 \times 10^{-8}$	$15.8 \times 10^6$	6.0	0.0005	0.45	37.0
1.4	$7.8 \times 10^{-8}$	$15.0 \times 10^6$	12.2	0.0010	0.85	11.1
1.5	$8.6 \times 10^{-8}$	$14.3 \times 10^6$	19.0	0.0014	1.05	8.3
1.6	$9.5 \times 10^{-8}$	$13.7 \times 10^6$	26.0	0.0008	1.17	16.5

For a set of values of P in column (1), column (2) is obtained from the values of  $A_3$  and E assumed. Column (3) follows from fig. 2, column (4) is obtained from equation (5.6). Interpolation in fig. 4 allows  $\mu_0$  to be obtained, and equation (3.10) and column (4) give the values of  $r_1$  in column (6). Then the values of  $K_1$  in column (7) are obtained from column (4), using the now known values of  $\mu_0$  and  $r_1$ . The values of  $K_1$  so obtained, as a function of P, give a minimum at about  $K_1=8.2$ . The correct pair of values of  $A_3$  and  $K_1$  will give this minimum at the same value of  $K_1$  as that corresponding to the value of  $A_3$  chosen. Although the calculations are not very accurate, unless more figures are retained than is the case in Table II.—which in turn would necessitate greater accuracy than that obtainable from fig. 4—by taking a number of pairs of values of  $A_3$  and  $K_1$  for which the corresponding minimum is obtained, it is not difficult to find sufficiently accurate values of  $A_3$  and  $K_1$ . It is found that these values are

$$\left. \begin{aligned} A_3 &= 3.5 \times 10^{26} \\ K_1 &= 9.0 \times 10^{-8} \end{aligned} \right\} \dots \dots \dots (7.2)$$

We now use equation (6.7) applied to the minimum of the curve of fig. 9 (b) to obtain  $\alpha \lambda_0 / K_1$ .

The minimum for the values of  $A_3$  and  $K_1$  in (7.2) occurs at a value of P at about 1.48 with  $I=2.7$  amps. We find that for this minimum

$\mu_0=0.00135$  and  $\frac{K_1 E^2 P^{-1} \mu_0 r_1^2}{4 T_\omega^2} = 18$ . From equation (3.9) we obtain  $T_0^2=32.3$ ;  $g(T_0^2)$  follows from fig. 7, and we obtain at once from (6.7)

$$\frac{\alpha \lambda_0}{K_1} = 36. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7.3)$$

Having obtained the constants in (7.2) and (7.3) the solution of the problem is complete. The experimental data on which the constants in (7.2) and (7.3) have been deduced, even apart from the lack of knowledge of the electrode losses, are really insufficient to ensure that the best values have been obtained; the constants should be calculated from several separate sets of data. They are, however, almost certainly sufficiently close to the truth to serve our present purpose.

### 8. Comparison with Experiment.

Although it is not expected that the theory given here will allow accurate calculations of the various parameters to be made, it is worth while seeing to what extent it may be expected to do so. We therefore make a few such tests.

In the first place, the constants we have found,  $K_1$  and  $\frac{\alpha \lambda_0}{K_1}$  allow us to find values for  $\lambda_0$  and  $\sigma_0$ . We have from (2.17) and (2.7)

$$9 \times 10^{-8} = K_1 = 0.0109 \frac{\lambda_0}{\sigma_0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.1)$$

and from (7.3) and (8.1)

$$\left. \begin{aligned} \sigma_0 &= 0.785 \text{ ergs cm.}^{-1} \text{ sec.}^{-1} \text{ } ^\circ \text{K}^{-1} \\ \lambda_0 &= 6.48 \times 10^{-6} \text{ cm.} \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.2)$$

So far as  $\sigma_0$  is concerned, from the measured value of  $7.75 \times 10^2$  at  $203^\circ \text{C}$ . and the relation  $\sigma = \sigma_0 T$ , we find  $\sigma_0 = 1.62$ . The temperature and pressure conditions under which our calculations have been made are such as would make any better agreement than this fortuitous. No measurements of conductivity have been made under anything like these conditions. The value of  $\lambda_0$  likewise cannot be compared satisfactorily with experimental values, partly because there do not seem to be any really reliable experimental results under any conditions, and partly because even if there were they would not be a very good check with the abnormal conditions in the high pressure arc; Elenbaas <sup>(2)</sup> uses the value  $2.17 \times 10^{-6}$  cm. All that can be said so far is that the values found check up roughly with what we should expect.

We now calculate  $T_0$  for the case considered by Elenbaas <sup>(1)</sup>, the parameters for which are given in (7.1). Thus from (5.6) we get  $\frac{K_1 E^2 P^{-1} \mu_0 r_1^2}{4 T_\omega^2} = 35$ ; and using this in (3.9), together with the value of  $T_1$

obtained from fig. 2, we easily obtain  $T=7260^\circ\text{K}$ . Elenbaas' experiment led him to the value of  $6000\pm 800^\circ\text{K}$ . There is no serious disagreement here.

Next, we attempt to calculate the curve of fig. 9 (a). The calculations are similar to those given in Table II. For two given values of  $E$ , namely 9.0 and 8.6 volts, we use the constants given in (7.2) and calculate the derived relation between  $P$  and  $K_1$ . The two appropriate values of  $P$  are those for which  $K_1=9\times 10^{-8}$ . From these values of  $P$  we can find the corresponding values in columns (3), (4) and (5),  $T_0$  and  $g(T_0^2)$  are obtained from equation (3.9) and fig. 7 respectively, and then the value of  $I$  from equation (6.7). The points so obtained corresponding to the two values of  $E$  are given as circles in fig. 9 (b). It is seen that the agreement is not good at the smaller values of  $I$ . It must be remembered, however, that the full line in fig. 9 (b) has been obtained from fig. 9 (a), assuming a constant voltage loss at the electrode of 16 volts. The two curves will coincide if the watts lost at the electrodes for large values of current is, in fact, proportional to the wattage, but falls off less rapidly than linearly at the lower wattages, having a value of some 30 watts at a tube wattage of 200. This is rather what would be expected and the value of the losses required seem about right. Even if this is the case, however, this does not provide a very good test of the theory. Since the constants have been estimated to fit certain parameters in the region of pressure of 1 to 2 atmospheres, and for values of  $E$  of about 8 volts, it would not be expected that any discrepancies would show up strongly in this region. A better test is to estimate a parameter for a tube with a considerably different pressure.

For this purpose we consider the commercial 125 watt quartz lamp. Such a lamp with

$$\left. \begin{array}{l} r_0=0.4\text{ cm.} \\ M=0.003\text{ gram.} \\ I=1\text{ amp.} \end{array} \right\} \dots\dots\dots (8.1)$$

has 125 volts across the arc. Again, to make the calculations we have to assume something about electrode losses. A total voltage drop at the two electrodes of 20 volts corresponding to a loss of 20 watts is about right for this type of lamp. Calculations were made in the following way:—A value of  $T_0$  was assumed, for example  $6400^\circ\text{K}$ ., and from equation (6.7) the corresponding value of  $\mu_0$  was calculated, namely 0.0050. A set of

values of  $\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_0^2}$  was then set out as in Table III., in which only a few values are given for illustration.

From the values in column (1)  $T_1$  is obtained from equation (3.9), then  $A_3 P^{3/2}/E^2$  from fig. 2. Column (5) is given by equation (5.6) and column (6) by equation (3.10). We then obtain from fig. 4, with the help of columns (1) and (3), a set of values of  $\mu_0$  which, since they cannot all be equal to the value of  $\mu_0$  corresponding to the value of  $T_0$  chosen, we call

$\mu_0'$ . The values of  $E^2$  in column (8) follow from column (3), and finally we get a new set of values of  $\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}$  in column (9) from the value of  $P$ ,  $r_1$ ,  $\mu_0$  and  $E^2$  in the four preceding columns. In general, there will be two rows in which the dashed value of  $\mu_0$  is equal to the true value, and also two different rows in which the dashed value in the last column is equal to the value  $\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}$  in the first column. There will be two values of the originally chosen  $T_0$  for which one of the rows where the correct value of  $\mu_0$  is obtained also has identical values of  $\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}$  in the first and last columns. One of these values of  $T_0$  corresponds to the operating point on the left of the minimum of the I, E characteristic with the appropriate value of IE (see fig. 9 (a)), and the other value of  $T_0$

TABLE III.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}$	$T_1$	$\frac{A_3 P^{3/2}}{E^2}$	$S\left(\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}\right)$	$P$	$r_1$	$\mu_0'$	$\frac{E^2}{(\text{volts})^2}$	$\left(\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}\right)$
12.2	5360	$2.38 \times 10^8$	0.252	4.86	0.127	0.0130	1570	33.8
19.0	4690	$3.80 \times 10^8$	0.234	5.23	0.232	0.0100	1100	58.0
22.0	4360	$4.90 \times 10^8$	0.226	5.43	0.269	0.0058	905	36.7
26.0	3870	$7.75 \times 10^8$	0.216	5.66	0.300	0.0014	608	7.3

$$T_0 = 6400 \quad \mu_0 = 0.0050.$$

will represent a similar point on the right of the minimum. The 125 watt lamp under consideration normally operates to the right of the minimum, so that we are interested only in the latter value of  $T_0$  and the corresponding values of the other discharge parameters. It is quite easy, by setting out the work systematically, to find the approximate required value of  $T_0$  which allows the calculations to reproduce the values of  $\mu_0$  and  $\frac{K_1 E^2 P^{-\frac{1}{2}} \mu_0 r_1^2}{4T_\omega^2}$  at the same time. The value of  $T_0$  was just below  $6400^\circ$  and the corresponding value of  $E$  was rather more than 28 volts. The experimental value, assuming the loss of 20 volts at the electrodes, is 35, so that the agreement is not particularly good. The discrepancy may be partly due to inaccuracies in the experimental data, but most of it is probably due to the inadequacy of the theory. The value of 40–45 volts loss at the electrodes, which would be necessary to give agreement with the calculated value, is out of the question for the type of lamp considered. The disagreement may be mainly due to the basic assumptions made in developing the energy balance equation, or it may be due to the device



adopted in section 3 for its solution, namely that assuming the energy dissipation uniform within the radius  $r_1$ . It is difficult without calculation to see how far this assumption would influence the value of  $E$  found from the equations. It would be too much to expect anyway, that accurate quantitative results could be obtained. It is clear, however, that qualitatively the mechanism of the discharge is adequately represented by the theory, and, in fact, the comparisons made with experimental results suggest that rather more than qualitative agreement is given.

It is hoped in a later paper to develop the theory further, together with its implications.

#### References.

- (1) W. Elenbaas, *Physica*, i. p. 211 (1934).
- (2) W. Elenbaas, *Physica*, i. p. 673 (1934).
- (3) W. Elenbaas, *Physica*, ii. p. 169 (1935).
- (4) W. Elenbaas, *Physica*, ii. p. 757 (1935).
- (5) W. Elenbaas, *Physica*, ii. p. 763 (1935).
- (6) F. Rössler, *Ann. de Phys.* xxxiv. p. 1 (1939).

---

#### XLIV. *New Contributions to Interferometry.*—Part VIII. *Interferometric Examination of Small Local Irregularities in Mica.*

By S. TOLANSKY, Ph.D., D.I.C., Manchester University\*.

[Received May 14, 1945.]

[Plates VII. & VIII.]

#### Abstract.

Two local irregularities in thin mica slips are studied by multiple beam interference methods using (a) monochromatic Fizeau fringes, (b) new white-light fringes called "fringes of equal chromatic order." The Fizeau fringe technique formerly employed has been developed so that high magnification (80 times) can be employed. The features studied extend over less than 1 millimetre. The fringes of equal chromatic order prove to be a powerful weapon for the interpretation of the features observed. One irregularity consists of a small surface hillock adjacent to a complex occlusion pocket. The hillock is interpreted as being a growth nucleus, and is shown to be built up in discontinuous steps. The hillock is about 2500 A.U. high and 34 steps have been measured in it. Of these, 56 per cent. are less than  $3\frac{1}{2}$  "molecules" and 85 per cent. less than  $7\frac{1}{2}$  "molecules" in height.

The second irregularity appears as a hillock flanked by two valleys. This is considered to be due to the inclusion of a small transparent foreign solid body during the laying down of the mica. The observed valleys are optical, arising from vacuoles associated with the foreign body. Two closely related possible mechanisms for the lay down of the mica

---

\* Communicated by the Author.

sheets over the foreign body are proposed. The definition obtained with the fringes representing this second feature supersedes all earlier observations.

---

### *Introduction.*

In Parts V. and VII. a description has been given of the application of monochromatic Fizeau fringes and of multiple beam white light fringes (named fringes of equal chromatic order) to the study of topographical features of mica. In Part VII. the detailed application to doubly silvered mica (mica silvered on both faces) was considered. It was shown that on cleavage, muscovite mica produces stepped areas of uniform thickness, and according to the sample these may be large or small.

Some samples show particularly large numbers of small cleavage areas (see, for example, Pl. IV. (c) in Part VII.). In addition to these, small roughly circular areas can be seen on Pl. VIII. (c), each less than 1 millimetre across, which exhibit circularly shaped rings resembling Newton's rings. These local features occur on several samples, and the present section is devoted to a detailed analysis of two such characteristic small areas.

The two localized irregularities discussed here have been studied with both multiple beam Fizeau fringes and with multiple beam fringes of equal chromatic order. Being on a small scale it was necessary to develop a fringe magnification technique. In the work formerly described the Fizeau fringes were mostly photographed directly by projection with a camera lens onto a plate, with a magnification of the order of four times. It was considered desirable to use magnifications up to about 80 times in the present studies. The fringes were therefore photographed with a simple microphotographic apparatus. The interference film was mounted on the stage of a polarizing microscope giving magnifications of 22, 78 and 350 with the objectives available, and the eyepiece replaced by a Leica microscope camera. The usual precautions as to collimation and illumination with monochromatic radiation (green mercury) were taken. The Nicol prisms on the microscope were convenient for the isolation of the separately polarized components.

With this arrangement exposure times, according to circumstances, varied from 2 minutes to 30 minutes. Despite the high magnification, remarkable fringe definition has been secured. Since there is no control over fringe width (such control is available in air-wedge topographical studies by increasing or decreasing the wedge angle), it was at first feared that the fringes might not tolerate high magnification. This proved groundless for the particular features studied.

The two specimens of mica examined were cleaved from the same thick piece of mica, and accidentally had nearly the same thickness, 1/136th and 1/140th millimetre. The silvering was better on one sample than on the other, leading to quite distinctly better fringes. This will be clear in the reproductions. In the regions of high reflecting coefficient

(approaching 0.90) a 1 per cent. change in reflecting power has a very considerable effect upon fringe definition, and in one of the examples studied the reflecting power attained was decidedly better than on the average. The improvement, although probably only a matter of 1 or 2 per cent., has a very marked influence on the fringe width. This is fortunate, for critical questions of definition and resolution are involved in the analysis of the effects observed. It may be mentioned that once silver has been put down upon a mica surface no attempt has been made to remove it for fear of affecting the surface during the removal process. If therefore a silvering is not quite up to standard, the specimen has always been discarded.

#### *Local Irregularity.—A.*

The Fizeau fringes of local irregularity (A) are shown in Pl. VII. (b) with a magnification of 22 times, and on Pl. VII. (a) with 78 times magnification. The mica was clamped between glass plates, and secondary fringes of low visibility can be seen. Observations were also made visually on the freely suspended mica.

There are two distinct parts to the irregularity, a marked circular hillock or valley, and a complex mass of fine scale fringes. The diameter of the hillock (it will be proved later that it is a hill) is about 0.7 millimetre, from which it can be seen from Pl. VII. (a) that some of the fine scale fringes within the complex region are separated by no more than  $1/125$  millimetre.

It is perfectly clear from the continuity of the fringes linking the two features, that they are in effect one. The photographs have been taken without a polarizing device and thus show the birefringence doubling. A particularly unusual feature about this is that the dispersion of the doublet rings is *greater* in the outer ring than in the other, which is quite the reverse of the usual Newton's rings type of fringe. This gives a clue to the structure which will be discussed later.

The fringes of equal chromatic order for the circular feature are shown in Pl. VII. (c), with an enlargement in Pl. VII. (d). The scales are such that the scale on Pl. VII. (b) nearly corresponds with that of Pl. VII. (c), and that on Pl. VII. (a) with that on Pl. VII. (d). Identification is simplified if it is noted that the short central strip in Pl. VII. (d) (with the fringes sloping down to the right) corresponds to the broad bright-dark-bright strip close to the top of Pl. VII. (a), just above the second ring.

To obtain these fringes of equal chromatic order an enlargement of five times was used in projecting the image of the interference film on the spectrograph slit. This was done for two reasons, (a) the feature is so small that it was difficult to project the required region on to the slit without some enlargement, (b) the region is so complex that definition is considerably improved by enlarging, for reasons given previously. The fringes of equal chromatic order, as far as was possible to adjust, represent a section running through the centre of the circular ring system. This requires adjustment of the image to a small fraction of a millimetre.

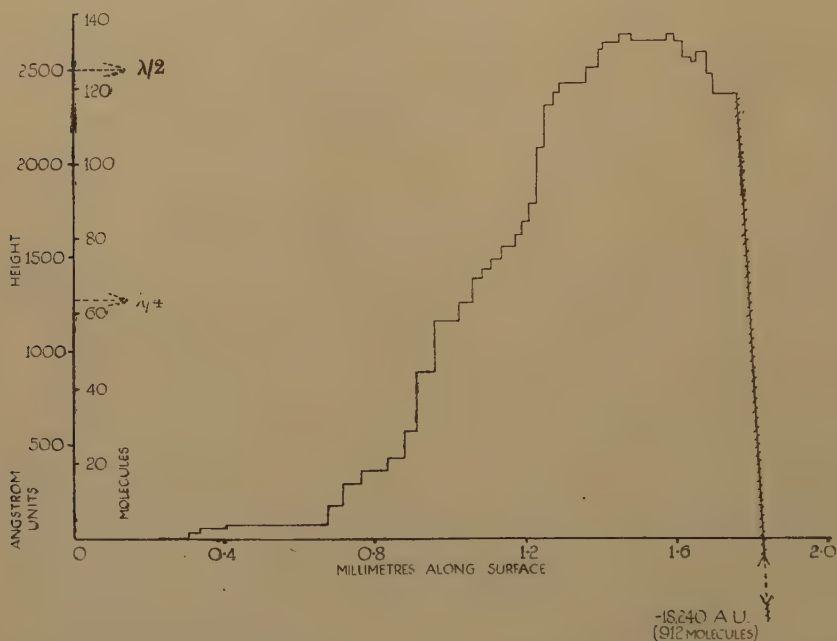


*Interpretation of the Fringes.*

From the Fizeau fringes alone it cannot be decided whether the circular rings belong to a hill or a valley. This point is settled by the fringes of equal chromatic order. These form well-marked bows concave to the violet, proving that the feature corresponds to an increase in optical path. It remains yet to show whether this is an internal or a surface feature. Fortunately this difficulty was easily resolved, for the whole complex could be quite easily seen on the surface of the mica with the aid of a lens. It is thus proved that the circular feature is a surface hillock.

The hillock must take on the general shape shown in fig. 1, for this is proved by the greater dispersion of the birefringence doubling in the

Fig. 1.



outer orders. This is, in fact, the unusual case of reverse Newton's rings, those which are formed by laying a concave surface upon a flat, a condition normally not usually realized experimentally. It is of interest to point out in this connection that the birefringence of the fringes of equal chromatic order is of no help at all with regard to this. It has been shown in Part VII. that the birefringence doubling with these latter fringes is independent of the mica thickness, and indeed it can be seen that the doubling remains uniform over the whole complex. On the other hand, with Fizeau fringes the dispersion is highly sensitive to the degree of parallelism, particularly when the surfaces are very nearly parallel, hence they show up the birefringence and also show a rapid increase in fringe width as the hillock descends gradually towards parallelism. Part of the third-order ring, very much broadened, can be seen in the horizontal bright upper strip on Pl. VII. (a).



*The Steps on the Hillock.*

There are two distinct features about the manner in which the hillock is formed. There are a few major steps and a large number of minor steps. It is necessary to separate the complex features of the two mica faces in order to identify those belonging to the hill. It seems clear from the way in which some of the large steps cut across the Fizeau fringe picture that they are probably to be attributed to the other side of the mica. This was indeed proved to be the case by the following method. Visual inspection with a lens enabled the side containing the hill feature to be recognized. The silver was removed from the other side. The side containing the hill feature (still silvered) was then matched against a silvered optical flat, and the true topographical surface distribution photographed with high magnification. The resulting picture was exceedingly complex. It has already been shown in earlier sections that even parallel-sided areas of mica show complex surface fringes owing to general surface distortion. Thus the hill feature is superposed upon a generally contorted surface, which makes recognition of individual features a matter of considerable difficulty. However, the irregular complex enables this to be done. By this means it was shown without ambiguity that the main long cleavage steps are indeed to be attributed to the other side, as was suspected. In calculating the height of the hill, and in determining the step formation, these major cleavage steps have therefore to be disregarded. The hill is therefore somewhat lower than it appears to be. The peak is actually about 2,700 A.U. high (the observed heights require to be corrected for the increase in the optical path due to the refractive index exceeding unity).

Measurements of the steps within the hill have been made and are shown in fig. 1. The ordinate scale is considerably greater than the horizontal scale, which gives the position on the mica. The hillock is, therefore, much exaggerated. The ordinates are given in angstrom units, also in waves of light for  $\lambda$  5,000, and also in "molecules" of mica, taking the molecule to be 20 A.U. in height.

It is clear that the hillock is built up from a sequence of discrete steps. The accuracy of the measurements is of the order of some 5 A.U. Although it cannot be decided with any certainty, there seems some evidence that a number of the steps are half integral in terms of molecules. The following table shows the values of the steps in A.U., in "molecules" (approximate) and the position of the leading edge, in millimetres. A length of about 1.7 millimetres is covered.

The most probable explanation of this phenomenon is that we are dealing here with a growth nucleus which has been interrupted from spreading over the crystal surface by the complex area to the right of it. This was almost certainly originally an occlusion. Suppose we postulate that successive sheets of mica have been laid down and have not spread far, it is easy to see that a stepped hill can be built up in the manner shown by the continuous lines in fig. 2. Growth will also begin elsewhere, and the sheets (dotted) will spread to meet the boundaries of the hill. The

dotted sheets, which are successively laid down later, will not coincide with the sheets within the hill, and in any case the binding strength will be weakened along the hill surface.

Suppose now, cleavage takes place, as represented by the thick line with arrows. One can expect that when this reaches the hillside it will pass over the hill, taking the path shown. (The cleavage path may possibly not pass over the summit but may break through and truncate the hill also.) Thus it follows that the steps on the cleaved hillside are actually the steps of the sheets of mica laid down during the build up of the crystal. Fig. 2 is schematic only.

Position, mm. $\times 10^{-3}$	229	308	334	405	676	714	765	835	883	912
Step, A.U. - -	0	27	31	14	104	112	71	68	149	327
Step, Mols. - -	0	$1\frac{1}{2}$	$1\frac{1}{2}$	1	5	$5\frac{1}{2}$	$3\frac{1}{2}$	$3\frac{1}{2}$	$7\frac{1}{2}$	$16\frac{1}{2}$

Position, mm. $\times 10^{-3}$	957	1028	1057	1079	1104	1135	1172	1191	1211	1230
Step, A.U. - -	261	97	129	48	51	70	63	71	197	195
Step, Mols. - -	13	5	$6\frac{1}{2}$	$2\frac{1}{2}$	$2\frac{1}{2}$	$3\frac{1}{2}$	3	$3\frac{1}{2}$	10	10

Position, mm. $\times 10^{-3}$	1250	1276	1288	1358	1393	1403	1449	1479	1574	1592
Step, A.U. - -	229	70	39	99	86	43	47	-40	+30	-32
Step, Mols. - -	$11\frac{1}{2}$	$3\frac{1}{2}$	2	5	4	2	$2\frac{1}{2}$	2	$1\frac{1}{2}$	$1\frac{1}{2}$

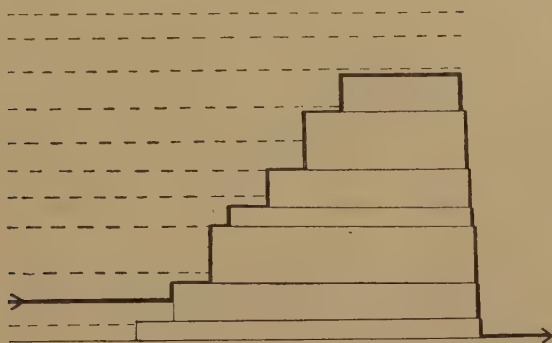
Position, mm. $\times 10^{-3}$	1614	1637	1652	1675	1690	1710	1816	1832
Step, A.U. - -	-90	-19	+48	-110	-111	—	-20600	+406
Step, Mols. -	$4\frac{1}{2}$	1	$2\frac{1}{2}$	$5\frac{1}{2}$	$5\frac{1}{2}$	—	1030	20

The values of these steps are worthy of comment. In all there are 34 steps in the hill feature. Of these, 19, *i. e.* 56 per cent., are below  $3\frac{1}{2}$  molecules in height, 29, *i. e.* some 85 per cent., below  $7\frac{1}{2}$  molecules high and only 5 steps (15 per cent.) between 10 and  $16\frac{1}{2}$  molecules in height. The steps are, therefore, truly molecular and small values are favoured. This may be associated with the rate of lay down.

It is clear from the reproductions that the hillock runs directly into the complex of fringes and both form a single feature. In the fringes of equal chromatic order there is a blur of undefinable structure, shown hatched, to the right of fig. 2. This blurred region is followed by two sets of intense fringes, which stand out brightly in the pattern of the fringes of equal chromatic order. The spacing of these fringes is much greater than of those forming the hill, hence the mica is much thinner here. The interpretation is that the hill climbs up steadily on the left of fig. 2 by a

series of small steps, and after crossing over, suddenly drops with extreme violence. Along a surface traverse of only 0.1 millimetre the hill falls through a total height of 20,600 angstrom units, *i. e.* over 1/500th millimetre. (It will be remembered that the whole mica thickness is only about three times this.) The drop, therefore, represents more than 1,000 molecules of mica. The sharp widely separated fringes in the valley are well defined, and show that the valley has a flat bottom and also exhibits a marked cleavage. Although not shown in fig. 2, it is clear from the reproduced fringes that the side of the depression opposite to the hill-drop rises by a series of steep jumps, and by the time the region shown at the bottom of Pl. VII. (c) is reached, the general mean surface level of the mica sheet has been regained.

Fig. 2.



It is quite clear that the complex of fringes represents a *pocket*, and this was originally either an air or a fluid pocket.

The drop to the right of the hillock is so violent that no distinctive fringes can be seen here, yet in spite of this it is quite evident that the hill feature represents a growth nucleus, the origin being associated with the occlusion pocket.

#### *Local Irregularity.—B.*

The local irregularity (B), to be discussed in this section, appears on a different piece of mica to that described above (irregularity A). The interference pictures illustrating this feature are shown in Pl. VIII. Fizeau fringes, with 22 times magnification, are shown on Pl. VIII. (a), and those with 78 times magnification on Pl. VIII. (b). The fringe definition is particularly good. Experimental conditions during the deposition of the silver mirrors were exceptionally good and the reflecting coefficients obtained were definitely higher than on the average. As a result the fringes given by this specimen are very sharp indeed. High precision in measurement is attainable and the fringes can be interpreted accordingly with confidence. The feature appears within a *highly uniform* field, blue to the unfiltered arc and quite black to the green filtered arc. Furthermore,



this uniform field is of some considerable extent, some square millimetres. This was the first hint that the feature may not be a simple surface defect at all.

Both Fizeau fringe pictures shown were taken *without* a polarizing device in the field, and the birefringence doubling is clearly seen. There are three separate features, a central contour and two satellites. In the satellites there are indications that the inner rings are broader than the outer, and this would suggest hollows on the surface or their optical equivalent. For the major central feature no decision on these lines is possible. As indicated formerly, the solution to this is given by the fringes of equal chromatic order, which are shown below in Pl. VIII. (c).

The definition in these fringes is superior to that obtained in any of the other fringes, and this enables critical decisions to be made. The section taken is somewhat to the left of the centre of the middle feature, and it is to be remembered that the hills and valleys do not stop at the ring boundaries but carry on, and extensions only appear visibly if the hill is high enough to enclose a complete second order, etc. It is quite clear that the upper and lower satellites are *valleys*, whilst the central feature is a *hill* standing up little more than one order above the surrounding flat region. The depths of the valleys shown are somewhat less than the maximum depths, because these valleys have not been centrally intersected by the slit.

#### *Interpretation.*

In interpreting this fringe distribution a number of alternative possibilities have been considered and rejected. In each of the rejected cases some physical feature has appeared improbable. The interpretation given below is considered to be a probable close approximation to the truth, hence it will only be necessary to indicate briefly the rejected alternatives.

A striking feature shown by the fringes of equal chromatic order is the difference in fringe smoothness upon the sides of the hillock as compared with the approach sides of the flanking valleys. The excellent definition renders this distinct even in the reproduction. The pictures in Pl. VIII. were taken with freely suspended mica (no glass supports) in order to eliminate any complications from secondary fringes. The purity of the Fizeau fringe pattern is evident, and all the details in the fringes of equal chromatic order are significant.

It is considered that the whole feature owes its origin to the inclusion of a small foreign body during the laying down of the mica crystal from solution. Such inclusions are of frequent occurrence in some micas and often of considerable thickness and opacity. The particular inclusion here is small, thin and transparent. It is no more than half a millimetre across and cannot be more than 1/150th millimetre in height. It is not unusual for so thin a section to be transparent. The occluded material is almost certainly a solid, not liquid, for the observed assymetry in the pattern is unlikely if a small liquid drop had been occluded. One might have expected then that surface tension forces for so small a droplet



would lead to a closer approach to a circular pattern. Further, one might have also expected a surrounding circular depression rather like a moat instead of the two dimples observed. Solids are so frequently found occluded in mica that the adoption of this view is quite consistent with experience.

Two separate mechanisms can be devised for the explanation of the observed fringe patterns, both closely related. The first mechanism is illustrated in fig. 3. It is necessary to assume that the occluded foreign body has a higher refractive index than the mica. In the laying down of the mica crystal from solution, one can picture the surface as covered with mother liquor except over the region on which the foreign body rests (experience shows that occlusions are frequently crystal). The laying down of the mica sheets continues over the surface in the manner illustrated. At some distance from the foreign body there is smooth deposition, and this is confirmed by the perfection of the fringes in the region near the scale in Pl. VIII. (c). As the mica sheets approach the foreign body, it is clear that a region will exist close to the edge wherein there will be either a shortage of mother liquor, or an exhaustion of dissolved material. This will depend upon the geometrical configuration of the foreign body. The lay down of the sheets will be interfered with on approaching the edge, and the situation shown in fig. 3 will arise.

Fig. 3.

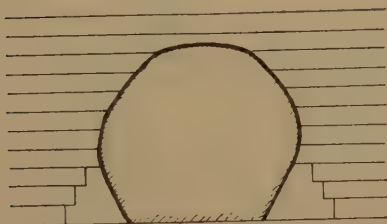
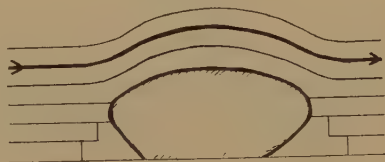


Fig. 4.



Deposition will continue after the foreign body is covered, and the actual mica surface is *not* distorted when a cleavage is carried out. The hillock and valleys are therefore "*optical hills and valleys*" only, owing their appearance to differences in refractive index within the slip of mica.

Calculation shows that this view is quite reasonable. The optical height of the hillock is some 3700 A.U. and the depth of the deeper valley some 2500 A.U. The mica is  $1/140$ th millimetre thick. Suppose we assume that the occluded body is half the thickness of the mica, *i. e.*  $1/280$ th millimetre thick, then the observed hillock will result if the refractive index of the occluded body is only greater than that of mica by 0.10, *i. e.* an index of 1.69 suffices. If the occlusion is still thicker the required refractive index is less. It cannot be less than 1.64 to fulfil requirements.

It will be seen that the valley appears because of the absence of solid material, either mica or foreign body, and there is no difficulty in obtaining suitable numerical conditions. A vacuole  $1/2500$ th millimetre thick suffices to produce the observed "*valley*."

The proposed mechanism explains also a number of the finer details. Steps can be seen on the "valley" sides in the interference picture, and this is as it should be if fig. 3 is correct. The mica is laid down in sheets from opposite sides, and it is quite possible that a discontinuity might take place on the peak of the foreign body where the sheets meet. Such a discontinuity can be observed in the interference picture. It is of some interest to point out that the X optic axis lies almost diagonally across the pattern in Pl. VIII, (b). If it be assumed that the growth sheets have spread laterally preferentially along this axial direction, then the formation of the "valleys" along this path might well be accounted for.

It should be noted that the occluded foreign body is higher on the left of the pictures than on the right. It can be considered to taper down towards the general level of the face of the mica on which it lies. For this reason there are two "valleys," on either side of the higher end.

Another mechanism can be proposed, and in this it is necessary to postulate the existence of a very thin foreign body, less than a wave in height, in fact the height to be of the order of the observed hillock. According to this view the mica sheets on the crystal surface approach the foreign body and lead to a vacuole at the edge, as before. Simultaneously mica is laid down on the surface of the foreign body and extends to meet the sheets spreading along the mica crystal surface. On meeting, a continuous mica sheath has now been formed and successive sheets are deposited over this, but now follow the contour of the occluded body. This leads to the formation of a true hillock, flanked by "optical" valleys. The hillock persists through succeeding layers and on cleavage will appear on the surface, as illustrated by the heavy line in fig. 4.

The height of this hillock is of the order of the height of the occluded body (small differences in refractive index introduce small modifications). As illustrated, the hillock height would approximate to half the height of the occluded body, the actual amount depending upon the shape.

Other mechanisms have been considered and rejected on various grounds. For example, a simple air or liquid pocket can hardly account for the observations. It should be possible to distinguish between the mechanisms of figs. 3 and 4 by examination of the mica surfaces separately. The feature is very small and, being superposed on the general surface twists and distortions, such observations were impossible to carry out.

The direction of the mica deposition layers changes in fig. 4, so that crystallographically fig. 3 might be preferable. To this must be added the fact that all other elements of mica surfaces yet examined are parallel to the second mica surface. On these general grounds the mechanism of fig. 3 is to be favoured.

My thanks are due to Prof. H. G. Cannon, F.R.S., of Manchester, for the loan of the micro-camera used in the above work.

# XLV. *A Photo-elastic Study of Stress Waves.*

By D. A. SENIOR, B.A., and A. A. WELLS, B.Sc.\*,  
Naval Construction Research Establishment.

[Received April 11, 1946.]

[Plates IX.—XII.]

## *Summary.*

A method has been devised whereby the stress waves set up by impact in simple models may be examined by photography of the photo-elastic fringes associated with them. For this purpose an Arditron discharge tube, giving single flashes of high intensity and duration less than 1 microsecond, is used in conjunction with a circuit which is closed by the pellet and which triggers the flash when a suitable time interval from the instant of impact has elapsed.

Fringe photographs are shown for a number of cases involving the impact of lead air-gun pellets with the edges of plates of photo-elastic material. In the materials tested the relationship between stress and birefringence appears to be of the same order with dynamic as with static stresses.

---

## PREVIOUS WORK.

### *The Photo-elastic Study of Dynamic Stresses.*

THE technique of photo-elasticity as applied to the analysis of plane stress distributions in loaded models has been developed by many investigators. In addition, work has been done with stresses due to impact by cinematography, and with cyclical variations of dynamic stresses by stroboscopic methods.

The cinematographic method of recording dynamic stresses has been carried out in two ways, firstly by Tuzi and Nisida from 1928 to 1936, who used continuously moving film exposed to a brightly illuminated slit superimposed on to the cross-section of a cantilever beam. Impact was applied to the end of the cantilever, thus causing bending stresses to be set up at the root. These stresses gave rise to sets of fringes whose distance apart varied with the magnitude of the bending moment. The record therefore consisted of a continuous set of equidistant lines which shifted normally at the instant of impact. With the strongest possible continuous light source and highest writing speed, however, no record of the detail of stress propagation was obtained.

The second method, that of straightforward cinematography at speeds up to 1000 or more frames a second has been carried out by a number of investigators. At this speed each frame shows a stress distribution

---

\* Communicated by J. F. Baker, O.B.E., M.A., Sc.D., M.Inst.C.E., M.I.Struct.E.



corresponding to static equilibrium, without any hint as to the manner in which stress propagation has taken place.

Stroboscopic recording of cyclical variations of dynamic stress has been carried out, both using a continuous light source in conjunction with a mechanical shutter, and by means of a stroboscopic neon discharge tube. Here again the recorded stress distribution is that corresponding to equilibrium throughout the body.

So far as is known, no record exists of fringe photographs of stress distributions having been obtained under kinematic conditions (*e. g.*, when a stress wave is travelling through a body). In order to obtain such fringe photographs, the exposure time must, in practice, be of the order of 1 microsecond and the instant at which the exposure occurs must be controllable with great precision. The source of illumination must fulfil this condition of short duration, and in addition be of high intensity.

#### *Spark and Flash Photography.*

Electrical sparks in air provide such a source of illumination and their use as a means of studying fast-moving subjects was suggested by Toepler in 1864. The method has since been developed by many workers and applied to ballistic photography, and to the photography of shock waves in gases and liquids, and other phenomena.

Electrical discharge through certain rare gas mixtures has recently been used to provide increased intensity of illumination at the expense of some increase in duration. The gaseous discharge tubes which have been developed possess vastly improved triggering characteristics compared with spark gaps, and the Arditron\* which has been used in the experiments described below is probably in this respect the most satisfactory tube so far developed.

#### *Theory of Stress Waves.*

##### *Stress Wave Velocity.*

The velocity of an extensional stress wave, travelling along a cylindrical bar, is partly given by the expression  $C_1 = \sqrt{E/\rho}$ , where  $E$  is the kinematic value of Young's Modulus for the material and  $\rho$  is the density. In general, the kinematic value of  $E$  is larger than the static value, especially for such materials as plastics where creep strain under static loading cannot be neglected.

For transparent plastic such as xylonite, perspex and C.R. 39, the basic stress-wave velocity is therefore of the order of 5,000 ft. per second.

For waves in two dimensions (thin sheet) and three dimensions, the velocities are modified to the values

$$C_2 = \sqrt{E/\rho(1-\nu^2)},$$

$$C_3 = \sqrt{\frac{E(1-\nu)}{(1-2\nu)(1+\nu)\rho}},$$

---

\* Developed by the Armament Research Department.



where  $\nu$  is Poisson's Ratio. For  $\nu=0.3$  the velocities are 5 per cent. and 16 per cent. greater for two and three dimensions respectively, so that the order of the values remains unchanged.

Photo-elastic examination is confined to waves of one and two dimensional plane stress. When impact takes the form of a concentrated force the wave fronts are respectively straight lines or approximate circles.

### *Peak Pressure Developed.*

Theories have been developed by Saint Venant, Boussinesque and others for the stresses set up by the longitudinal impact of rods of equal cross-sectional area. Compression waves are set up in both rods. These travel from the point of impact with velocities given by  $\sqrt{E_1/\rho_1}$ ,  $\sqrt{E_2/\rho_2}$ , the suffixes 1 and 2 referring respectively to the striking and struck rods. The maximum pressure at the interface occurs at impact, exists until the wave reflected from the remote end of one of the rods arrives at the junction, and is given by

$$P=V \left[ \frac{\sqrt{E_1\rho_1} \times \sqrt{E_2\rho_2}}{\sqrt{E_1\rho_1} + \sqrt{E_2\rho_2}} \right],$$

where  $V$  is the relative velocity of impact.

If the cross-sectional areas of the rods are unequal, the wave front in the rod of larger cross-section will exhibit a reduced stress at points remote from the junction; a reasonable approximation is  $P/A$ , where  $A$  is the cross-sectional area.

For rods of equal cross-section, if the striking rod is shorter than the struck rod and if  $E_1\rho_1 > E_2\rho_2$ , the stress wave imparted to the struck rod takes the form of a stepped wave, the stress decreasing in the ratio  $\frac{\sqrt{E_1\rho_1} - \sqrt{E_2\rho_2}}{\sqrt{E_1\rho_1} + \sqrt{E_2\rho_2}}$  at intervals of length  $\frac{\sqrt{E_2\rho_2}}{\sqrt{E_1\rho_1}} \cdot 2l$ , where  $l$  is the length

of the shorter rod. If  $\frac{E_1\rho_1}{E_2\rho_2} \rightarrow \infty$  the peak stress is given by  $V\sqrt{E_2\rho_2}$ , and the stress behind the wave front decreases exponentially according to the law

$$f = V\sqrt{E_2\rho_2} \cdot e^{-\frac{\sqrt{E_2\rho_2}}{\rho_1 l} t}.$$

Since in practice the surfaces where impact occurs are imperfect, stress will initially be propagated from the points where contact first occurs, and the resultant stress wave will not exhibit the infinite stress gradient indicated above.

For the case of a lead pellet striking the end of a xylonite specimen of the same cross-section, the stress wave imparted may be expected to exhibit a stress gradually increasing to a maximum (the stress gradient being greater for higher velocities of impact) the subsequent decrease not being exponential since the pellet is not rigid but flows on impact.

If it is assumed that the pellet behaves as a fluid flowing out at right angles to the direction of flight, and if the velocity of the pellet is greater than that of plastic waves in the material of which it is composed, the initial pressure at the interface is  $\rho_1 V^2$ , and this is maintained for a time  $l/V$ . For the pellets used in the series of experiments here described, the ratio of this pressure to that obtained by assuming the lead to be rigid is  $\frac{\rho_1 V^2}{V\sqrt{E_2\rho_2}} = \frac{\rho_1 V}{\rho_2 C_1} = 0.93$ , *i.e.* the pressures are of the same order.

Under the particular experimental conditions, it therefore seems reasonable to take the value  $V\sqrt{E_2\rho_2}$  as a basis for estimating the photo-elastic sensitivity of the material under dynamic conditions.

When the xylonite specimen consists of a bar of greater cross-section than the pellet, the stress wave imparted to the specimen will at points remote from the junction be similar in form to that described above.

When the specimen consists of a large sheet, the stress pattern at any instant before reflection will involve stresses which are constant along lines roughly forming semi-circles whose centre is at the point of impact. Thus the principal stress difference is constant along these lines. Close to the edge where impact occurs, the stress pattern will be modified by the presence of tension waves travelling outwards along the edge.

The velocity of the stress front will be  $\sqrt{E/\rho(1-\nu^2)}$ , but, as the wave is two-dimensional, its form will vary as it travels outwards. Attenuation will also occur. Thus the stress difference at points in the specimen on the line of flight of the pellet may or may not reveal the peak stress depending upon the distance which the stress front has travelled.

#### *Experimental Details.*

Since the stress wave travels through the specimen with a velocity of approximately 5,000 ft. per second, and since the linear dimensions of the specimen are necessarily limited by those of the polariscope to a few inches, a "triggering" signal representing the instant of initiation of the stress wave must be produced accurately to within a few microseconds. In a suitably chosen specimen, a stress gradient sufficient to produce up to 5 isochromatic fringes per inch can be produced by even a low velocity of impact (*e.g.* 20 ft. per second), provided that good contact is obtained. It is, however, preferable to use a higher velocity (*e.g.* 400–500 ft. per second) and a photo-elastic material of but medium sensitivity since triggering which depends upon contact is then much more precise. Impact loading is therefore carried out by means of a 0.22 in. bore air pistol, the lead pellets being filed as shown in fig. 1*a* so as to present a flat surface to the specimen and thus provide the most accurate triggering signal, and produce a stress wave of the required form.

The signal itself is derived from two strips of tinfoil which are cemented to the specimen as shown in fig. 1*b*. Electrical contact between these strips occurs when the pellet bridges the gap between them.

The light source consists of an Arditron tube through which is discharged a 0.5 microfarad condenser charged to a potential of 7,500 volts. The Arditron is mounted directly on the condenser in order to minimise

the self-inductance of the discharge circuit and therefore the duration of the flash. Results indicate that under the conditions of the experiment the effective photographic duration of the flash is between 0.25 and 0.5 microsecond.

Fig. 1.

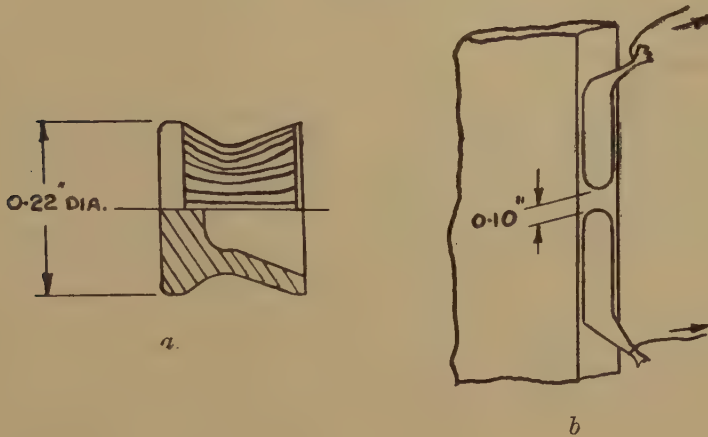


Fig. 2.

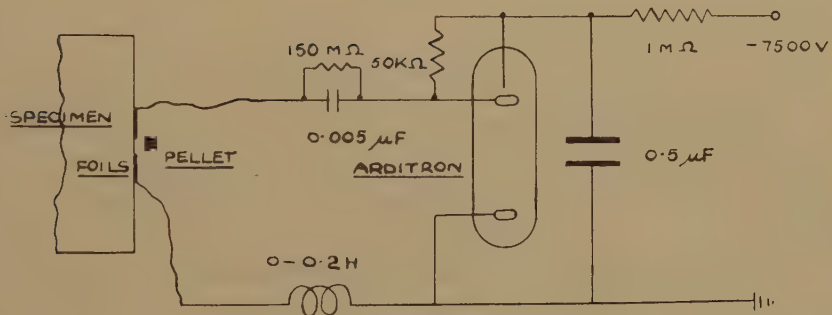
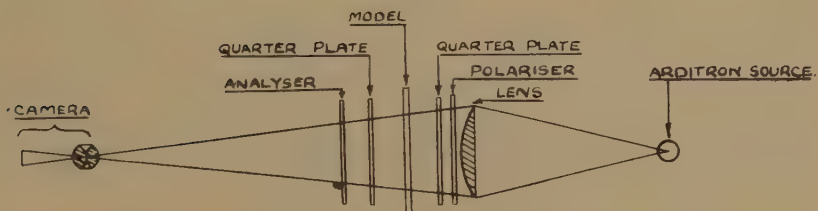


Fig. 3.



The flash is timed relative to the instant of initiation of the stress wave by means of the circuit of fig. 2. This circuit was found more consistent, and simpler to operate, than several electronic circuits which were also tried.

With the optical arrangement shown in fig. 3 and a camera lens aperture of  $f/4.5$ , sufficient illumination is provided for high contrast process plates

to be used. The advantage of these plates is that they are insensitive to light of wave-lengths greater than 5,000 A.u., and fringe photographs are therefore obtained reasonably free from chromatic effects. Further improvement arises from the facts that the glass components of the polariscope cut off wave-lengths less than about 4,000 A.u., and a considerable portion of the light output from the Arditron tube is of wave-length between 4,300 and 4,500 A.u., in which region the sensitivity of the plate is a maximum.

The general arrangement of the polariscope is shown in fig. 3. The quarter-wave plates eliminate isoclinic fringes, thereby secure greater clarity, and simplify interpretation of the isochromatic fringe pattern. The particular quarter-wave plates used are designed for operation at 5,461 A.u., corresponding to the mercury green line. Even when operated with wave-lengths between 4,500 and 5,000 A.u., however, they effectively eliminate isoclinics. The number of lens elements in the polariscope is cut down to a minimum for the sake of simplicity. Rays pass through the specimen obliquely, but the confusion in the fringe pattern caused by this is not serious.

### *Results.*

Series of fringe photographs taken at different intervals of time after impact of the pellet are shown for four types of target.

(a) (See Plates IX. and IX. *a.*) The pellet is impinging on to the middle of the long edge of a transparent xylonite plate  $6\frac{1}{2}$  in. by  $4\frac{1}{2}$  in. by  $3/16$  in. The pressure wave appears to expand radially with uniform velocity in all directions except along the edge of the plate, where there are tension waves. Interaction of the pressure and tension waves produces a complex stress distribution. Isotropic points are formed and maintained, moving out with the same velocity as the compression and tension wave fronts.

(b) (See Plates X. and X. *a.*) The same target as at (a) is used but in expanding the wave front passes a small circular hole drilled through the plate. Reflections which modify the oncoming pressure wave are caused by the free edges of the hole. When the peak of the pressure wave passes the hole stress concentrations at its periphery are seen to be formed and the stress distribution in the immediate vicinity of the hole becomes similar to that produced by a uniform static axial compression in the direction of wave travel. As far as can be seen the stress concentration factor at the hole, at the instant when the pressure peak passes it, is approximately 1.8. The stress concentration factor in this case is defined as:

$$\text{Stress concentration factor} = \frac{\text{Peak hole boundary stress}}{\text{Peak pressure when hole is absent}}.$$

(c) (See Plates XI. and XI. *a.*) The plate is decreased from  $4\frac{1}{2}$  in. to 1 in. wide, so that an early reflection of the radially expanding pressure wave occurs. The pressure wave, at each successive reflection, gives rise to tension waves which are propagated along the free edges of the bar.



(d) (See Plates XI. and XI. a.) The pellet impinges on to the end of a long rod of rectangular section 1 in. by  $\frac{3}{16}$  in. A pressure wave is set up which becomes uniformly distributed transversely at points remote from the struck end of the bar. Near the struck end the stress distribution is complicated by reflection and also by the bending set up by asymmetrical impact. The small mass of the impinging pellet relative to the bar causes the pressure wave to die away in a time of the same order as that required to build up.

The calculated value of the peak stress produced by the pellet impinging on to the rod at 450 fps. is given by  $V\sqrt{E\rho}$  as 36,000 lbs. per square inch. Distributing this over the cross-section of the rod reduces it to 7,300 lbs. per square inch, or  $7,300/4$  lbs. per square inch per fringe. The dynamic photo-elastic coefficient is therefore 8 Brewsters. The value measured for statically applied stresses is 12.8 Brewsters.

### Conclusions.

It is believed that this method of photo-elastic stress analysis of phenomena associated with stress propagation will be of use in the solution of many problems concerned with ballistics and the design of structural members subjected to impact.

More information would be obtained if a multiple flash camera were used. For the type of problem dealt with in this paper, a camera giving up to 20 exposures at intervals of 5 or 10 microseconds would be suitable. A system is at present being developed whereby a pair of independent photographs is obtained with a measured time interval of 0 to 50 microseconds.

The experimental results indicate that the photo-elastic constant for dynamic stresses in xylonite is approximately 8 Brewsters, compared with 13 Brewsters for statically applied stresses. The discrepancy may be due to inaccuracy in calculation of the peak stress in the stress waves investigated.

The authors' thanks are due to the Admiralty for permission to publish this paper.

## XLVI. On the Cause of the Red Shifts in the Spectra of the Extra-galactic Nebulæ.

By G. J. WHITROW\*.

[Received April 1, 1946.]

### 1. Introduction.

IN a recent lecture, Hubble<sup>(1)</sup> has summarized his views on the structure of the universe, as revealed by the observational data. Briefly, the main characteristics of the region so far observed are :—

(a) the system of extra-galactic nebulæ is isotropic and shows no signs of thinning out towards the observable frontier ;

\* Communicated by the Author.

(b) if it is assumed that, on the average, the nebulae are relatively stationary, the large scale distribution is sensibly homogeneous ;

(c) the radial velocities of the members of the local group indicate that this group maintains its dimensions ;

(d) the spectra of the more distant nebulae are shifted towards the red, the displacement increasing as the distance increases.

Originally, the red shifts were interpreted as Doppler shifts, due to radial recession from the observer. It was thought that any other *known* cause of spectral displacement would give rise to phenomena which would be conspicuous and, in fact, are not observed. As Hubble remarked in his Rhodes Lectures <sup>(2)</sup>, "We may state with some confidence that red-shifts are the familiar velocity-shifts, or else they represent some unrecognized principle of nature. We cannot assume that our knowledge of physical principles is yet complete ; nevertheless, we should not replace a known, familiar principle by an *ad hoc* explanation, unless we are forced to that step by the actual observations." However, Hubble claims that the Doppler interpretation of the red shifts leads to values for the age \* and extent of the universe, which seem to be dubiously small. On the other hand, the assumption that the whole observable region maintains its dimensions, in similar fashion to the local group, leads "to the very simple conception of a sensibly infinite homogeneous universe, of which the observable region is an insignificant sample." <sup>(3)</sup> Hubble, therefore, concludes that "a choice is presented, as once before in the days of Copernicus, between a strangely small, finite universe and a sensibly infinite universe plus a new principle of nature."

Although, as has been pointed out elsewhere <sup>(4)</sup>, Hubble is mistaken in arguing that a density-distribution increasing radially outwards necessarily implies a favoured position for the observer, the agreement he claims to have obtained between observation and theory, in the case of a static Euclidean uniformly dense model, is striking. Any such comparison depends, of course, on the choice of initial assumptions, and different investigators have arrived at conflicting conclusions. For example, McVittie <sup>(5)</sup> has argued that the Doppler interpretation of the red shifts points to a hyperbolic universe of infinite extent, and, more recently, Heckmann <sup>(6)</sup> has criticized Hubble's procedure.

On the theoretical side, Milne and the author <sup>(7)</sup> maintain that the same model can be described as static or expanding, according to the choice of time-scale and associated length-scale. At first sight it would appear that Hubble's results belie this possibility, but this is not so, because in his analysis Hubble adopts, in effect, the *same* scale throughout. Working on orthodox lines, he uses the clock and measuring rod of "classical" physics. If he did not, his objection, that the age of the universe and its extent appear to be dubiously small when red shifts are interpreted as Doppler effects, would lose its crucial significance. In fact, Hubble's

---

\* Less than  $10^9$  years.

analysis, if correct, would seem to show that, on his initial assumptions, e. g., adoption of the "classical" scales of time and space measurement, estimation of nebular distances on the hypothesis of average uniform absolute brightness, etc., the model in most satisfactory agreement with the present range of observations is static. For an expanding model to be in satisfactory agreement, the initial assumptions must be appropriately and consistently modified.

The principal *theoretical* objection to adopting a static model is that there then appears to be no explanation for the existence of "red shifts." Hence, on his assumptions, Hubble concludes that they are inexplicable. In the present note I propose to consider this question further. Actually, it will be found sufficient to consider only Hubble's implicit assumptions concerning the measurement of time. By paying close attention to this, it will be found that the existence of "red shifts" in a static universe is, in fact, a direct consequence of certain general properties which it is plausible to assign to the universe, whether finite or infinite. Thus, there is no need to regard this phenomenon as being in any way surprising or mysterious, nor is it incompatible with a static model, e. g., the one favoured by Hubble.

## 2. The Scale of Time in Einstein-Newtonian Physics.

The concept of a steady state depends on the adoption of a definite time-scale. The *point d'appui* of Einstein's theory of relativity was his realization that the epochs assigned by an observer to *distant events* are conventional. He assumed, however, that the time-scale used by each observer for recording local events was the absolute time,  $\tau$ , of Newtonian mechanics\*. This time "flows uniformly" for all observers, but its rate of flow and origin of measurement are arbitrary; in other words, the equations of Einstein-Newtonian physics are independent of a linear transformation of time-scale,  $T = a\tau + b$ . Thus,  $\tau$  itself does not appear explicitly in the laws of physics, because the constant  $b$  is indeterminate. The constant  $a$ , however, can be conventionally chosen in terms of a given periodic phenomenon, e. g., the earth's rotation (suitably "corrected"), and hence time-differences and, in particular, the differential element  $d\tau$ , have a meaning; the latter, of course, appears explicitly in classical formulæ.

The *point d'appui* of Milne's theory of relativity is his explicit recognition of the conventional character of the epochs assigned by an observer to *local events*. In principle, the observer can adopt any time-scale,  $T = f(\tau)$ , where  $f$  is no longer restricted to be linear but can be any monotonic increasing function, the rate of increase of which, however, can tend to zero as  $\tau \rightarrow \pm \infty$ . If, following the method of Special Relativity (and of radar), distances are determined by epoch measurements

---

\* Of course, in mathematical theory General Relativity admits all possible regratuations of the observer's time-scale, but in all practical applications Einstein assumed that there is a unique *natural* time-scale associated with each observer.



in light, or other electromagnetic, signalling experiments, the description of a particular system as steady or non-steady depends on the choice of clock  $f(\tau)$ . Thus, the interpretation of red shifts as due to recession is meaningless unless a definite scale of time is specified.

Confining attention to a static universe, we will assume that the time-scale associated with this description is  $\tau$ . It will be found that there are phenomena pointing to the co-existence of another "uniform" time-scale in Nature; or, if in accordance with the normal custom of physicists, we prefer to use only one scale of time, we shall find that some of the parameters in the fundamental laws of physics must vary with change of epoch. To use a convenient term, coined by Sir Edmund Whittaker<sup>(9)</sup>, this situation can be expressed as a "postulate of impotence": *it is impossible to choose a scale of time and an associated scale of length so that all the fundamental parameters of physics are independent of change of epoch.*

### 3. *The Apparent Luminosity of a Distant Nebula in an Infinite Universe.*

In 1826, Olbers suggested that an infinity of stars (similar to the sun) would either cause infinite sky-brightness or at least would make each region of the sky appear to be as bright as the sun. Hence, if there are an infinite number of uniformly radiating nebulae, it follows either that light is modified in some way during its transmission through intergalactic space, or else that atomic oscillators do not vibrate uniformly in  $\tau$ -time; for, if neither of these effects occurred, the total apparent brightness of the system would be as Olbers suggested. Hence, as has been pointed out previously<sup>(10)</sup>, red shifts, whatever their cause, are an essential feature of an infinite world-model of uniformly radiating nebulae, because they provide the necessary dimming factor.

There are three possible ways in which red-shifts might be caused in a static universe:—

(a) absorption, scattering, etc., of light in inter-nebular space might occur in such a way that the total observed light intensity at any point is finite and of the observed order of magnitude; or

(b) the frequencies of light quanta might slow down during long periods of transmission in inter-nebular space; or

(c) the uniform time of *all* light quanta might differ from  $\tau$ -time.

Of these three hypotheses, the last is the only one which could be submitted to a laboratory test: moreover, it is the only one in general agreement with other phenomena to be considered in the following section of this note.

Let us suppose then that the uniform time of atomic vibrators is  $T=f(\tau)$ , where  $f$  is some function which is non-linear. Consider a photon of frequency  $n$ , which is emitted by a distant nebula at epoch  $\tau$ . From the definition of the  $T$ -scale, it follows that the frequency ( $n$ ) in this scale



will be unaltered during transmission. The corresponding frequency in the  $\tau$ -scale at epoch  $\tau$  will be  $\nu(\tau)$ , where

$$\nu(\tau)d\tau = n dT.$$

Hence

$$n = \frac{\nu(\tau) d\tau}{dT} = \frac{\nu(\tau)}{f'(\tau)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Thus, in the  $\tau$ -scale the apparent frequency of a given photon at time  $\tau$  will be proportional to  $f'(\tau)$ . The intensity of light from a distant source will tend to zero (a necessary condition for finite sky brightness in an infinite universe of uniformly radiating nebulæ), if, and only if, the apparent frequency of each photon emitted tends to zero. If the universe is infinite, the light received now from the most distant nebulæ was emitted when  $\tau$  was arbitrarily large and negative. Thus, for finite sky brightness, it is necessary that  $f'(\tau) \rightarrow 0$  as  $\tau \rightarrow -\infty$ . Consequently, on our hypothesis, the  $\tau$ -rate of atomic vibrators in the remote past was infinitely slow, and hence the spectra of the most distant nebulæ in an infinite universe will appear to be shifted indefinitely towards the red.

#### 4. The Uniform Time of Radio-active Decay.

A more powerful argument for the existence in Nature of a uniform time-scale other than  $\tau$  arises from considering the phenomenon of radio-activity. The argument applies irrespectively of whether the universe is infinite or finite.

The Rutherford-Soddy law,

$$\frac{dN}{N} = -\lambda dT,$$

expresses the probability that  $dN$  atoms of a deposit of  $N$  atoms of a particular radio-active element will decay in an interval of time  $dT$ . For elements such as uranium, thorium, etc., it has been found that, within the range of experiment, the values of  $\lambda$  are independent of all macroscopic conditions, e. g., temperature, pressure, etc. It thus appears that these values are determined solely by conditions within the atoms themselves. The assumption that these values are independent of epoch can be regarded as the *definition* of the uniform time  $T$  of radio-active decay<sup>(11)</sup>.

Consider the integral form of the decay law, viz.,

$$N(T) = N_0 \exp \{-\lambda(T - T_0)\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where  $N_0$  is the number of atoms present at an arbitrarily chosen epoch,  $T = T_0$ . Either this law holds for all possible values of  $T$  or else we must, assume that decay began suddenly at a particular epoch. Suppose that  $T$  is Einstein-Newtonian time,  $a\tau + b$ , which extends from  $-\infty$  to  $+\infty$ . It follows that  $N$  was arbitrarily large in the remote past, unless decay began suddenly at a particular epoch. Hence, in a finite universe, decay, according to the law (2), cannot have occurred when  $\tau$  was arbitrarily

large and negative. On the other hand, in an infinite universe, decay, according to this law, would imply that the original source of each given small deposit of atoms of any radio-active element was extended over an indefinitely large region of space. In view of the existence of numerous independent radio-active deposits, this situation does not appear to be physically plausible. Thus, both in an infinite universe and, *a fortiori*, in a finite universe, it would appear either that radio-active decay of a given deposit began at a definite epoch, or else that the uniform time of radio-activity differs from Einstein-Newtonian time in such a way that, as  $\tau \rightarrow -\infty$ ,  $N$  is finite and hence  $dN/d\tau \rightarrow 0$ .

In view of the most recent development in atomic physics, the possibility of the initiation of decay of a naturally radio-active element at a definite epoch (subsequent to "creation") must be examined. If, in a static universe, the uniform time of radio-active decay were  $\tau$ , there would appear to be, on the average, the same probability at all epochs for some radio-activity to be initiated naturally. We have no evidence that this is so. On the contrary, the special conditions necessary for the initiation of natural radio-active decay probably occurred only in the remote past. The assumption that the uniform time of radio-activity differs from  $\tau$ -time automatically implies that conditions in the static universe change with lapse of time. Hence, we again conclude that radio-active decay is non-uniform in  $\tau$ -time and that  $N$  is finite for  $\tau$  "large" and negative. Although the argument is not so sharp in this case, it would appear that a "good approximation" is still obtained by regarding  $N$  as a function of  $\tau$ , which would remain finite if it were extrapolated to arbitrarily large negative values of  $\tau$ .

### 5. The Uniform Time of Atomic Processes in General.

We now consider, in detail, the restrictions which must be placed on the time-scale,  $T=f(\tau)$ , defined by the Rutherford-Soddy law. First, as  $\tau \rightarrow -\infty$ ,  $N(T)$  must be finite, and hence  $T$  must tend to a finite value. Furthermore,  $T$  must be a monotonic increasing function\* of  $\tau$ . Therefore, the value of  $T$ , as  $\tau \rightarrow -\infty$ , must be its initial value  $T_0$ , say. Now consider what happens as  $\tau \rightarrow +\infty$ .  $N(T)$  must either tend (a) to a finite value, or (b) to zero. In the former case,  $T$  tends to a finite value  $T_1$ , say; in the latter,  $T \rightarrow +\infty$ . Moreover, since  $dN/d\tau \rightarrow 0$ , as  $T \rightarrow T_0$ , it follows that  $f'(\tau) \rightarrow 0$  as  $\tau \rightarrow -\infty$ . Also, as  $\tau \rightarrow +\infty$ , either  $N \rightarrow 0$  or else  $f'(\tau) \rightarrow 0$ , and so, in either case,  $dN/d\tau \rightarrow 0$ . Thus, the rate of decay in Einstein-Newtonian time was arbitrarily slow in the remote past and will become arbitrarily slow again in the distant future. It follows that there is at least one epoch at which this rate attains a maximum value. These epochs depend on the form of  $f(\tau)$  and also on the value of  $\lambda$ , and so will

---

\* Also  $dT/dt$  must be a monotonic increasing function of  $t$ ; for, if it were not, there would exist singular epochs at which  $d^2T/dt^2$  would be zero. Similarly, for higher derivatives.

not be the same for all radio-active elements. They are given by the appropriate roots of the equation

$$f''(\tau) = \lambda [f'(\tau)]^2.$$

If, following the method of Special Relativity and of Kinematic Relativity, distances are calculated by multiplying appropriate time intervals by  $c$  (the velocity of light), the time-interval on the  $\tau$ -scale corresponding to the distance between two typical nebulæ, is a time-invariant; since they are at a fixed distance apart if their proper motions are neglected. The situation is different if we employ the  $T$ -scale, defined by the law of radio-active decay, and the associated length-scale, obtained by multiplying the appropriate  $T$ -intervals by  $c$ . Since  $T \rightarrow T_0$  as  $\tau \rightarrow -\infty$ , it follows that in the remote past a fixed interval on the  $\tau$ -scale corresponded to an arbitrarily small interval on the  $T$ -scale. Hence, according to the  $T$ -scale, the whole universe was originally compressed into an arbitrarily small volume. (The volume on the  $T$ -scale of any sub-system, e. g., a nebula, was also arbitrarily small originally.) Thus, according to this metrical scale, the universe originally expanded from a point-singularity, but without further specification of  $T$  it is not possible to say whether this expansion continues indefinitely or ultimately slows down, or even becomes eventually a contraction.

As previously stated, the parameter  $\lambda$  appears to be independent of macroscopic conditions. Since, according to the  $\tau$ -scale, the general distribution of matter is independent of epoch, it follows that, like  $\lambda$ , the variation of  $T$  with  $\tau$  must be determined solely by atomic phenomena. Thus, the uniform time of radio-active decay is essentially an "atomic" scale of time. Until there is reason to suspect the contrary, it is natural to assume that this time-scale is the same for all radio-active elements and is the uniform time of all atomic processes. In particular, it will be assumed to be the uniform time of radiative processes, such as the emission of photons from stars and nebulæ. Thus, as has already been shown for an infinite universe, the spectra of the more distant nebulæ will appear to be shifted, in general, towards the red, even if the universe is finite.

## 6. The Two Time-Scales of Kinematic Relativity.

So far, I have considered the most general form of  $f(\tau)$  consistent with the fundamental postulates. In practice, it is to be expected that  $f(\tau)$  is unique, except for choice of time origin and unit. Of the two possibilities, referred to in the first paragraph of section (5), attention will be confined to (b), as there is no reason to suppose that a given fraction of a deposit of atoms of a particular radio-active element will remain unchanged for all time. By considering possible forms of  $f(\tau)$  of this general type, it appears that the only elementary analytical function of this form is

$$T - T_0 = \alpha e^{\tau/\beta}, \quad . . . . . (3)$$







I pass now to the experimental determination of  $t_0$ . It is readily proved that, according to the  $t$ -scale given by (6), the fundamental particles of the universe, static on the  $\tau$ -scale, recede from each other with uniform velocities, the whole system occupying an arbitrarily small volume as  $t \rightarrow -t_0$ . Thus, on the  $t$ -scale, the parameter  $t_0$  is the present "age of the universe." This is determined by correlating the observed distances of the nearer nebulæ with their observed spectral shifts by the formula

$$t_0 = \frac{\Delta}{V} = \frac{\Delta/c}{\delta\nu/\nu}, \quad \dots \dots \dots (7)$$

where  $\Delta$  is the distance\* and  $\nu/(\nu - \delta\nu)$  is the observed red-shift ratio, which is the same at all epochs for a particular nebula if uniform atomic time is given by (6). The observations indicate that  $t_0$  is approximately  $1.9 \times 10^9$  years.

It follows that there are two equally valid alternative interpretations of  $t_0$ , viz. :—

(a) If the unit of  $\tau$ -time is fixed, e. g., the period of the earth's rotation round the sun, suitably corrected, then the distance  $\Delta$  of a given nebula will be constant, on the average, and hence  $t_0$  will be "a constant of Nature." Meanwhile, the associated unit of  $t$ -time, viz., the number of standard atomic vibrations in a unit of  $\tau$ -time, will increase as  $\tau$  increases, the ratio of the  $t$ -unit to the  $\tau$ -unit being given by the relation

$$\frac{t\text{-unit}}{\tau\text{-unit}} = e^{\tau/t_0} = 1 + t/t_0. \quad \dots \dots \dots (8)$$

(b) Alternatively, if the unit of  $t$ -time is fixed, e. g., the period of a specified number of standard atomic vibrations, then the distance  $\Delta$  of a given nebula varies proportionately to  $t$ , and  $t_0$  is the age of the system at the instant of observation. In this case the associated unit of  $\tau$ -time, viz., the number of rotations of the earth round the sun in a unit of  $t$ -time, decreases as  $t$  increases, according to the law

$$\frac{\tau\text{-unit}}{t\text{-unit}} = \frac{1}{1 + t/t_0} = e^{-\tau/t_0}. \quad \dots \dots \dots (9)$$

## 7. Some Consequences of the Interplay of the Two Time-Scales.

The system of units associated with the  $\tau$ -scale appears to be appropriate to ordinary and to celestial mechanics. Hence, on this scale of measurement, the ideal C.G.S. metre rod should be of invariant length. On the other hand, if the wave-length of a standard atomic emission or absorption line is invariant on the  $t$ -scale system, then, in comparison with the C.G.S. metre-rod, it will vary as  $t_0/(t + t_0)$ . Consequently, at the present time, the wave-length of any standard spectral line, e. g., the red line of cadmium, will be shrinking by about one part in  $2 \times 10^7$  per century.

---

\* For the nearer nebulæ the  $\tau$ -scale and  $t$ -scale distances are the same, to a first approximation, at the present epoch.

At present the laboratory determination of this particular wave-length,  $6438.4696 \times 10^{-10}$  metres at 15 deg. C. and normal atmosphere, is accurate to within about  $\pm 0.002 \times 10^{-10}$  metres. With improvement in technique, it should be possible in the next fifty years \* to test experimentally the validity of the assumption that all atomic vibrators keep  $t$ -time, particularly if full use is made of statistical methods.

To conclude this note, brief reference will now be made to three other consequences of the suggested inter-play of the two time-scales.

First, from equation (1) it follows that

$$\frac{\nu(\tau_1)}{f'(\tau_1)} = \frac{\nu(\tau_0)}{f'(\tau_0)},$$

where  $\tau_1$  refers to the instant of emission of a photon from a distant nebula and  $\tau_0$  to the present instant of reception. Hence, the law of red shifts will be given by the relation

$$\frac{\text{observed frequency}}{\text{standard frequency}} = \frac{f'(\tau_1)}{f'(\tau_0)} = e^{-(\tau_0 - \tau_1)/t_0} = e^{-\Lambda/t_0},$$

where  $\Lambda$  is the length, in  $\tau$ -measure, of the path described by the photon. This relation can be written in the form

$$\frac{\delta\lambda}{\lambda} = \frac{\Lambda}{ct_0} + \frac{1}{2} \left( \frac{\Lambda}{ct_0} \right)^2 + \dots, \quad (10)$$

where  $\lambda$  denotes the wave-length of a standard spectral line. To the present limits of observation, the ratio of the second term on the right to the first rises to not more than about six per cent. It is doubtful if the uncertainties in the estimation of  $\Lambda$ ,  $\delta\lambda/\lambda$  and  $t_0$  permit us to distinguish, at present, between this relation and the one favoured by Hubble, namely,

$$\frac{\delta\lambda}{\lambda} = \frac{\Lambda}{ct_0}. \quad (11)$$

Second, in the case of an infinite universe it can be verified, on the lines originally indicated by Milne, that the total apparent brightness is finite.

Finally, Haldane has shown that, if  $t$ -time is the uniform time of radioactive decay, then, since

$$\frac{dN}{dt} = -\lambda N = -\lambda N_0 e^{-\lambda(t+t_0)},$$

it follows that

$$\frac{dN}{d\tau} = -\lambda N_0 \left( \frac{t+t_0}{t_0} \right) e^{-\lambda(t+t_0)}.$$

---

\* *Vide* Haldane, 'American Scientist,' xxxiii. p. 131 (1945).

Hence, both as  $t \rightarrow -t_0$  and as  $t \rightarrow +\infty$ , the Einstein-Newtonian time-rate of radio-active decay tends to zero, and there is a maximum at the epoch  $t$ , where

$$t+t_0=1/\lambda.$$

For uranium ( $\text{U}^{238}$ ) and thorium ( $\text{Th}^{232}$ ),  $t$ -scale intervals of  $4.4 \times 10^9$  years and  $2 \times 10^{10}$  years, respectively, must still elapse before these maxima are attained. Haldane <sup>(15)</sup> has indicated some geological evidence in favour of the view that the  $\tau$ -scale rates of decay of uranium and thorium are increasing.

### 8. Summary.

It has been suggested previously <sup>(16)</sup> that two different scales of "uniform time" operate in Nature, the red shifts in the spectra of the extra-galactic nebulæ being due to the fact that when a distant light source is observed the out-of-stepness of the two scales becomes apparent. In the present note an attempt has been made to account for this duality. It has been shown that "improbable" consequences would result, if the uniform time-scale of atomic phenomena were the same as the time-scale of classical mechanics, which can be associated with a static universe, e. g. Hubble's static universe of uniformly radiating nebulæ. A possible qualitative explanation is thus suggested for the existence of the red shifts in the spectra of these nebulæ. Further discussion on quantitative lines, based on the assumption of mathematical simplicity, leads to the isolation of Milne's atomic time-scale  $t$ , with its implication that, on the average, the spectral shift of a particular nebula, as observed from any other given nebula, is the same at all epochs. The resulting formula, correlating predicted red-shift ratio with assigned distance as estimated by Hubble, using the "classical"  $\tau$ -scale, agrees satisfactorily with the present range of observations. The possibility of a laboratory test to confirm the existence of the  $t$ -scale in Nature has been indicated.

### References.

- (1) Hubble, Amer. Sci. xxx. pp. 99-115 (1942).
- (2) Hubble, 'The Observational Approach to Cosmology,' p. 26 (1937).
- (3) Hubble, Amer. Sci., *loc. cit.*
- (4) Milne and Whitrow, *Zeit. f. Astrophys.*, xv. pp. 263-98 (1938).
- (5) McVittie, Proc. Phys. Soc. li. pp. 529-37 (1939).
- (6) Heckmann, "Theorien der Kosmologie," *Fortschritte der Astronomie*, ii. § 17 (1942).
- (7) Milne and Whitrow, *op. cit.*
- (8) Milne, 'Relativity, Gravitation and World-Structure,' § 26 (1935).
- (9) Whittaker, Proc. Roy. Soc. Edin. lxi. pp. 160-75 (1942).
- (10) Milne and Whitrow, *op. cit.*
- (11) Whitrow, 'Nature,' cxliv. p. 306 (1939).
- (12) Milne, Proc. Roy. Soc. (A), clviii. p. 327 (1939).
- (13) Milne, *Zeit. f. Astrophys.* vi. p. 95 (1933).
- (14) Haldane, 'Nature,' cliii. p. 555 (1944).
- (15) Haldane, Amer. Sci. xxxiii. pp. 129-45 (1945).
- (16) Milne and Whitrow, *op. cit.*





remaining symbols. Further, it has been found that, except for ammonia, water, the alcohols and acetic acid, each power-index in equations (1) to (6) differs little in value from liquid to liquid, similarly to  $n_3$ ,  $n$  and  $p$  in the equations of the Verschafield, Ferguson and Kleeman types

$$D-d=D_0(1-T/T_c)^{n_3}, \quad \sigma=\sigma_0(1-T/T_c)^n, \quad \sigma=\sigma_1(D-d)^p.$$

TABLE I.  
Constants of Latent Heat-Temperature Relations.

Substance	$T_c$	$m_0$	$\lambda_0$	$n_0$	$l_0$
Pentane .....	470.3	0.42749	121.84	0.40035	131.56
Isopentane .....	460.9	0.42792	116.18	0.39368	124.42
Hexane .....	507.9	0.42461	115.62	0.39836	124.22
Diisopropyl .....	501.5	0.43616	110.11	0.40198	117.65
Heptane .....	540.0	0.41436	111.19	0.38293	117.97
Octane .....	569.3	0.41583	106.64	0.39095	113.81
Diisobutyl .....	549.9	0.40109	95.92	0.37581	102.71
Benzene .....	561.6	0.41826	129.65	0.38288	138.50
Carbon dioxide .....	304.5	0.46712	138.60	0.45251	155.35
Carbon disulphide .....	546.2	0.48165	116.56	0.40395	126.43
Chloroform .....	533.1	0.41462	83.77	0.33156	84.81
Carbon tetrachloride .....	556.3	0.41172	62.43	0.37929	67.12
Fluorobenzene .....	559.7	0.42379	108.77	0.38989	116.01
Chlorobenzene .....	632.3	0.43341	106.87	0.37639	110.59
Bromobenzene .....	670.1	0.45193	81.83	0.38561	83.88
Iodobenzene .....	721.1	0.46184	68.29	0.38468	69.28
Ethyl ether .....	466.9	0.42383	120.58	0.39690	129.58
Acetone .....	510.6	0.41125	180.41	0.33231	182.21
Methyl formate .....	487.1	0.44422	158.51	0.41767	171.01
Ethyl formate .....	508.4	0.44305	138.47	0.41493	148.47
Propyl formate .....	538.0	0.41918	122.43	0.39388	131.39
Methyl acetate .....	506.8	0.43124	141.55	0.40595	151.88
Ethyl acetate .....	523.2	0.42981	126.69	0.40744	136.11
Propyl acetate .....	549.3	0.42694	118.46	0.40092	126.20
Methyl propionate .....	530.5	0.42452	125.67	0.39707	134.07
Ethyl propionate .....	546.0	0.42550	116.12	0.40234	124.31
Methyl butyrate .....	554.4	0.41794	114.07	0.38950	121.24
Methyl isobutyrate .....	546.7	0.43699	113.49	0.40460	120.11
Tin tetrachloride .....	591.8	0.43826	44.15	0.39882	46.62
Sulphur dioxide .....	429.1	0.40802	122.86	0.38327	132.75
Nitrous oxide .....	311.9	0.41525	108.25	0.41977	132.66
Ammonia .....	404.1	0.40380	428.02	0.33301	438.46
Water .....	643.1	0.37476	689.68	0.31820	708.66
Methyl alcohol .....	513.1	0.40889	380.19	0.38256	399.68
Ethyl alcohol .....	516.2	0.43205	307.43	0.40254	319.93
Propyl alcohol .....	536.8	0.51135	274.12	0.47900	284.37
Acetic acid .....	594.8	0.35150	136.87	0.33952	149.75

The evaluation of the constants of the equations (1) to (6) for each liquid has been carried out by equating the logarithms of the expressions on opposite sides of the equation, then substituting in turn in the loga-

rithmic equation so obtained each recorded experimental value of the latent heat and its corresponding reduced temperature, or orthobaric densities, or surface tension, and finally applying the method of least squares to the resulting numerical simultaneous equations.

## 2. Latent Heat and Temperature.

The experimental values employed for calculating the constants given in Table I. include those for all the liquids whose internal and total latent heats over a range of temperature greater than  $45^\circ$  are recorded in the literature; they will be found collected in a paper by Mills <sup>(7)</sup> and are largely based on the work of Young <sup>(8)</sup>. For most of the liquids the range of temperature is more than  $200^\circ$ , but for carbon dioxide, nitrous oxide

TABLE II.  
Latent Heat and Temperature.

Temp. ° C.	Tin tetrachloride				Water			
	$\lambda$ calc.	$\lambda$ obs.	$l$ calc.	$l$ obs.	$\lambda$ calc.	$\lambda$ obs.	$l$ calc.	$l$ obs.
0	33.66	33.30	36.42	35.38	560.6	565.0	594.4	595.8
20	—	—	—	—	549.1	552.5	583.9	584.8
40	—	—	—	—	537.1	539.2	573.1	573.5
60	—	—	—	—	524.7	525.4	561.8	561.8
80	—	—	—	—	511.7	511.4	550.0	549.8
100	28.54	29.01	31.34	31.76	498.2	497.1	537.6	537.5
120	27.37	27.70	30.17	30.54	484.0	481.0	524.6	523.0
140	26.12	26.21	28.92	29.12	469.1	465.0	510.9	508.5
160	24.80	24.75	27.58	27.69	453.4	449.0	496.3	494.2
180	23.38	23.33	26.14	26.29	436.7	433.6	480.7	479.7
200	21.83	21.59	24.56	24.51	418.9	418.2	464.1	465.3
220	20.14	19.98	22.82	22.82	399.7	402.1	445.9	449.5
240	18.24	18.15	20.85	20.86	378.8	381.7	426.1	429.5
260	16.04	16.02	18.55	18.50	355.8	355.7	404.0	402.5
270	—	—	—	—	343.4	344.7	392.0	390.3
280	13.36	13.48	15.71	15.60	—	—	—	—

and ammonia is less than  $100^\circ$ , so that the values of the constants given in Table I. for these three liquefied gases must be regarded as less accurate than those for the other liquids. The only values of latent heats that have been excluded from the calculations are those given for temperatures within  $10^\circ$  of the critical temperature of the substance, and for acetic acid the values for temperatures below  $172^\circ \text{C.}$ , where according to the spectroscopic observations of Gillete and Daniels <sup>(9)</sup> the dimeric form of the acid is present in the vapour.

Using the values of the constants in Table I., the applicability of the equations to the calculation of latent heats, in calories per gram, is shown for tin tetrachloride and water in Table II.

Table I. shows that the values of the indices  $m_0$  and  $n_0$  are fairly constant for normal liquids. Excluding the indices for the molecularly associated liquids, ammonia, water, the alcohols and acetic acid, the mean value of  $m_0$ , weighted for each liquid proportionally to the range of temperature over which its latent heats are recorded, is 0.42852, and similarly that of  $n_0$  is 0.39297. Of these, the former is represented closely by  $3/7$ , but not by  $1/2$  as proposed by Tyrer<sup>(6)</sup>, and the latter by  $7/18$ , or, less well, by  $2/5$  proposed by Winter<sup>(5)</sup> or by  $1/3$  suggested by Thiesen<sup>(6)</sup>, so that the approximate general equations for a non-associated liquid are

$$\lambda = a_0(1 - T/T_c)^{3/7}, \quad l = b_0(1 - T/T_c)^{7/18},$$

where the values of the constants  $a_0$  and  $b_0$  may be obtained directly by substituting the known values of  $T_c$ ,  $T$  and either  $\lambda$  or  $l$  in these two equations and taking the means of the values thus obtained. In Table III. the observed values of the latent heats of octane and benzene are compared with those calculated from the general equations.

TABLE III.  
Latent Heats calculated from the General Equations.

Temp. ° C.	Octane $a_0=109.21$ ; $b_0=113.38$ ; $T_c=569.3$				Benzene $a_0=131.83$ ; $b_0=139.86$ ; $T_c=561.6$			
	$\lambda$ calc.	$\lambda$ obs.	$l$ calc.	$l$ obs.	$\lambda$ calc.	$\lambda$ obs.	$l$ calc.	$l$ obs.
0	82.55	84.71	87.95	89.46	99.09	99.16	107.94	106.11
80	—	—	—	—	86.22	86.70	95.13	95.45
100	—	—	—	—	82.57	82.37	91.48	91.41
120	66.07	65.10	71.86	71.43	78.69	77.39	87.57	86.58
140	62.74	61.78	68.57	68.28	74.55	73.45	83.38	82.82
160	59.16	58.17	65.01	64.75	70.06	69.48	78.81	78.94
180	55.27	54.36	61.12	60.91	65.16	65.21	73.80	74.62
200	50.97	50.21	56.79	56.61	59.72	59.75	68.17	68.81
220	46.13	45.90	51.87	52.03	53.51	53.76	61.71	62.24
240	40.48	40.37	46.08	45.97	46.15	46.53	53.95	54.11
260	33.53	34.25	38.83	39.14	36.74	37.55	43.88	43.82
280	23.76	24.67	28.40	28.26	—	—	—	—

### 3. Latent Heat and Orthobaric Density.

The results of the application of equations (3) and (4) are listed in Table IV., in which are included all liquids for which values of orthobaric densities and latent heats at various temperatures are available. The temperature range for which the constants of each liquid have been evaluated is the same as in the preceding section, and the same values of the latent heats have been used, the only exception in this respect being water, for which, in order to give the disturbing effect of maximum

density a wide berth, the range has been limited to 80–270° C. The densities are given in grams per cubic centimetre. The constants for mercury, which are based on the vapour pressure and density tables<sup>(10)</sup>, give the values of its latent heats at any temperature between 0° C. and the boiling point 356·7° C., with a maximum error of only 0·6 per

TABLE IV.  
Constants of Latent Heat-Orthobaric Density Relations.

Substance	$m_1$	$\lambda_1$	$n_1$	$l_1$
Pentane .....	1·3195	147·76	1·2369	157·70
Isopentane .....	1·3481	143·67	1·2410	151·35
Hexane .....	1·3283	137·87	1·2472	146·63
Diisopropyl .....	1·3635	130·59	1·2580	137·80
Heptane .....	1·3377	132·51	1·2372	138·81
Octane .....	1·3008	122·79	1·2243	130·06
Diisobutyl .....	1·2739	111·10	1·1946	117·95
Benzene .....	1·3150	112·01	1·2047	121·20
Carbon dioxide .....	1·3565	62·19	1·3141	71·47
Carbon tetrachloride .....	1·2768	25·07	1·1772	28·96
Fluorobenzene .....	1·3439	76·42	1·2372	83·86
Chlorobenzene .....	1·4131	70·52	1·2279	77·08
Bromobenzene .....	1·4796	34·85	1·2626	40·49
Iodobenzene .....	1·4899	21·89	1·2409	26·86
Ethyl ether.....	1·2526	118·60	1·1739	127·65
Acetone.....	1·3690	171·81	1·1063	175·16
Methyl formate .....	1·3660	108·15	1·2819	119·18
Ethyl formate .....	1·3696	104·48	1·2836	114·09
Propyl formate .....	1·3068	99·02	1·2287	107·68
Methyl acetate .....	1·3454	105·55	1·2676	115·27
Ethyl acetate .....	1·3323	100·76	1·2639	109·59
Propyl acetate .....	1·3478	99·14	1·2665	106·81
Methyl propionate .....	1·3384	99·52	1·2527	107·82
Ethyl propionate.....	1·3193	96·22	1·2484	104·11
Methyl butyrate .....	1·3257	95·08	1·2363	102·34
Methyl isobutyrate .....	1·3783	93·81	1·2772	100·73
Tin tetrachloride .....	1·3564	10·99	1·2350	13·15
Nitrous oxide .....	1·1132	56·93	1·1246	69·27
Mercury .....	1·6067	1·08	0·7563	10·29
Water .....	1·3776	522·61	1·1933	562·49
Methyl alcohol .....	1·4792	373·66	1·3860	393·60
Ethyl alcohol .....	1·5308	301·16	1·4289	314·20
Propyl alcohol .....	1·7097	255·42	1·6037	266·43
Acetic acid .....	1·0692	98·11	1·0354	108·81

cent. ; in comparing the constants for mercury with those for the other liquids, it has to be borne in mind that the former relate to a region remote from the critical temperature.

The observed values of the latent heats, and those calculated from the orthobaric densities and the constants for chlorobenzene and water, are given in Table V.



The mean value of  $m_1$ , weighted for each liquid proportionally to the range of temperature and excluding the constants for mercury, water, the alcohols and acetic acid, is 1.3441, and that for  $n_1$  is 1.2372, the former being about 4/3 and the latter 5/4, so that for a normal liquid, approximately,

$$\lambda = a_1(D-d)^{4/3}, \quad l = b_1(D-d)^{5/4}.$$

The results obtained by the application of these general equations to the data for pentane are shown in Table VI.; the values employed for

TABLE V.  
Latent Heat and Orthobaric Density.

Temp. °C.	D—d	$\lambda_{\text{calc.}}$	$\lambda_{\text{obs.}}$	$l_{\text{calc.}}$	$l_{\text{obs.}}$
Chlorobenzene					
0	1.12780	83.58	85.06	89.35	89.89
140	0.96798	67.35	66.48	74.06	73.36
160	0.94122	64.73	64.14	71.56	71.26
180	0.91217	61.93	61.67	68.85	68.96
200	0.88044	58.90	58.50	65.92	65.83
220	0.84575	55.65	55.55	62.76	62.89
240	0.80560	51.96	52.25	59.11	59.49
260	0.75988	47.84	48.17	55.02	55.15
270	0.73419	45.57	45.80	52.74	52.56
Water					
100	0.95778	492.5	497.1	534.3	537.5
120	0.94228	481.5	481.0	524.0	523.0
140	0.92443	469.0	465.0	512.1	508.5
160	0.90424	454.9	449.0	498.8	494.2
180	0.88144	439.2	433.6	483.9	479.7
200	0.85494	421.1	418.2	466.5	465.3
220	0.82538	401.2	402.1	447.4	449.5
240	0.79224	379.2	381.7	426.0	429.5
260	0.75527	355.0	355.7	402.4	402.5
270	0.73691	343.2	344.7	390.8	390.3

$a_1$  and  $b_1$  were found by substituting the experimental values of D,  $d$ , and either  $\lambda$  or  $l$  in the equations and taking the means of the values so obtained.

On introducing the molecular weight, M, equations (3) and (4) become respectively

$$\frac{M\lambda^{1/m_1}}{D-d} = M\lambda_1^{1/m_1} = [A],$$

$$\frac{Ml^{1/n_1}}{D-d} = Ml_1^{1/n_1} = [L],$$

where the quantity  $[A]$  may be termed the true lyoparachor and  $[L]$  the normal lyoparachor of the liquid. At temperatures below the boiling

point,  $d$  is practically negligible, and for the non-associated substances  $m_1$  has nearly the same value, so that at such temperatures the ratio of the true lyoparachors of two such liquids is the ratio of their molecular volumes under conditions in which their internal latent heats are the

TABLE VI.  
Application of General Equations.

Pentane.

$$a_1=149.25; \quad b_1=159.21.$$

Temp. ° C.	$D-d$	$\lambda_{\text{calc.}}$	$\lambda_{\text{obs.}}$	$l_{\text{calc.}}$	$l_{\text{obs.}}$
0	0.64459	83.10	85.85	91.95	93.36
40	0.60284	76.00	76.16	84.57	84.31
60	0.57898	72.02	71.69	80.41	80.07
80	0.55228	67.63	66.86	75.80	75.33
100	0.52144	62.64	61.57	70.55	69.94
120	0.48570	56.98	56.33	64.56	64.48
140	0.44010	49.96	49.08	57.07	56.58
160	0.38030	41.12	40.89	47.55	47.42
180	0.29320	29.07	30.02	34.35	35.01

same. The normal lyoparachor gives similarly a relative measure of the molecular volumes of non-associated liquids when their total latent heats of vaporisation are equal.

#### 4. Latent Heat and Surface Tension.

Table VII. includes all liquids for which the data necessary to calculate directly the values of the constants in equations (5) and (6) are available.

TABLE VII.  
Constants of Latent Heat-Surface Tension Relations.

Substance	$m'$	$\lambda'$	$n'$	$l'$
Benzene .....	0.33998	30.268	0.31122	36.569
Carbon tetrachloride .....	0.33498	15.217	0.30860	18.285
Chlorobenzene .....	0.35298	23.607	0.30667	29.788
Ethyl ether .....	0.34132	29.969	0.31964	35.184
Methyl formate .....	0.36536	32.440	0.34265	38.518
Ethyl acetate .....	0.34894	29.234	0.33078	33.898
Mercury .....	0.62025	1.593	0.29197	12.365
Methyl alcohol .....	0.37586	79.811	0.36212	90.298
Ethyl alcohol .....	0.42127	56.367	0.36614	70.653
Acetic acid .....	0.23173	44.068	0.21485	51.261

For each liquid the temperature range and values of latent heat employed in the calculations are the same as in the previous sections; the values

of the surface tension in dynes per centimetre are those obtained by application of the Ferguson-Kennedy constants<sup>(2)</sup>, but for mercury the measurements of surface tension from 0° to over 300° made by Didenko and Pokrovsky<sup>(11)</sup> have been used, while for the alcohols and acetic acid the values of surface tension taken are the experimental as corrected by Sugden<sup>(12)</sup>, their temperature ranges being those given by him.

The values of the latent heats given by the constants for carbon tetrachloride and ethyl alcohol are compared with the experimental in Table VIII.

TABLE VIII.  
Latent Heat and Surface Tension.

Temp. ° C.	$\sigma$	$\lambda_{\text{calc.}}$	$\lambda_{\text{obs.}}$	$l_{\text{calc.}}$	$l_{\text{obs.}}$
Carbon tetrachloride					
0	29.49	47.27	48.35	51.95	51.87
80	19.62	41.24	41.64	45.82	46.00
100	17.27	39.52	39.64	44.05	44.15
120	14.98	37.68	37.48	42.15	42.08
140	12.76	35.71	35.27	40.12	39.92
160	10.60	33.56	33.28	37.89	37.95
180	8.53	31.20	30.83	35.44	35.40
200	6.55	28.56	28.22	32.65	32.61
220	4.67	25.49	25.35	29.42	29.45
240	2.93	21.80	21.91	25.47	25.56
260	1.36	16.88	17.15	20.11	20.07
Ethyl alcohol					
20	22.27	208.3	208.0	220.1	220.6
80	17.97	190.3	191.6	203.5	206.4
110	14.49	—	—	188.0	190.3
140	11.34	156.8	155.7	171.9	171.1
170	7.85	—	—	150.2	148.4
200	4.26	103.8	104.0	120.1	116.6
230	0.95	—	—	69.3	70.6

The mean value of  $m'$ , weighted according to temperature range for the liquids exclusive of mercury, the alcohols and acetic acid, is 0.3468, that is, about 7/20, and that of  $n'$  is 0.3187, that is, about 5/16. The general equations therefore are

$$\lambda = a' \sigma^{7/20}, \quad l = b' \sigma^{5/16},$$

where  $a'$  and  $b'$  are constants; the applicability of these equations to the calculation of the latent heats of ethyl ether is shown by the results recorded in Table IX.

That the mean values given above for  $m'$  and  $n'$  hold well generally for normal liquids can be confirmed, for, besides the foregoing, two other independent methods of evaluating these indices are available. First,

on eliminating  $1 - T/T_c$  between Ferguson's equation and each of the equations (1) and (2), it is found that

$$m' = m_0/n, \quad n' = n_0/n.$$

Secondly, elimination of  $D - d$  between the equation of Kleeman and each of the equations (3) and (4) gives

$$m' = m_1/p, \quad n' = n_1/p.$$

Now the values of  $n$  and  $p$  for eighteen of the liquids listed in Tables I. and IV. are given by Ferguson and Kennedy <sup>(2)</sup>, and on combining their values for the most recent and extensive measurements with the values of  $m_0$  and  $n_0$  in Table I. and of  $m_1$  and  $n_1$  in Table IV., it is found that the mean value obtained for  $m'$  by applying both of these methods to the eighteen liquids is 0.343, and that for  $n'$  is 0.316, in agreement with

TABLE IX.

Application of General Equations.

Ethyl ether.

$$a' = 29.485; \quad b' = 35.679.$$

Temp. ° C.	$\sigma$	$\lambda_{\text{calc.}}$	$\lambda_{\text{obs.}}$	$l_{\text{calc.}}$	$l_{\text{obs.}}$
0	19.81	83.85	85.41	90.72	92.52
20	17.32	80.00	80.04	87.00	87.54
40	14.88	75.86	75.02	82.96	82.83
60	12.51	71.39	70.40	78.58	78.44
80	10.23	66.54	65.40	73.79	73.50
100	8.05	61.18	60.40	68.47	68.42
120	5.98	55.14	54.53	62.39	62.24
140	4.03	48.02	48.13	55.15	55.33
160	2.27	39.28	39.81	46.10	46.07
180	0.74	26.54	27.36	32.48	31.87

0.3468 and 0.3187 respectively, already here given and obtained by the method of least squares on the six liquids that have been experimentally investigated over the widest range of temperature.

### 5. Viscosity Relations.

Equations connecting the latent heat, orthobaric density or surface tension of a liquid with its viscosity can be obtained by combining equations (1), (2), the Verschaffelt equation and the Ferguson equation separately with the relation

$$\eta - \eta_c = (\eta_0 - \eta_c)(1 - T/T_c)^r, \quad \dots \dots \dots (10)$$

where  $\eta_0$ ,  $\eta_c$  and  $r$  are specific constants for the liquid. For instance, equations (2) and (10) yield the relation

$$\eta = \eta_c + \frac{(\eta_0 - \eta_c)}{l_0^{r/n_0}} \cdot l^{r/n_0},$$



and similarly, equation (10) and the Ferguson relation give the expression

$$\eta = \eta_c + \frac{(\eta_0 - \eta_c)}{\sigma_0^{r/n}} \cdot \sigma^{r/n}.$$

Such equations can be reduced to numerical form by substituting the values of the constants given in the preceding sections or the communications of previous workers <sup>(1), (2), (3)</sup>. Thus, for ethyl acetate the viscosity in poises is given by the equations

$$\eta = 0.00087 + 1.661 \times 10^{-14} l^{5.685}, \quad \dots \dots (11)$$

$$\eta = 0.00087 + 8.302 \times 10^{-6} \sigma^{1.880}, \quad \dots \dots (12)$$

as shown in Table X.

TABLE X.

Latent Heat, Surface Tension and Viscosity of Ethyl Acetate.

$$n = 1.232, \quad \sigma_0 = 66.90;$$

$$r = 2.3161, \quad \eta_c = 0.0008699, \quad \eta_0 - \eta_c = 0.022435;$$

$$n_0 = 0.40744, \quad l_0 = 136.11.$$

Temp. °C.	$l$	$\sigma$	$\eta_{\text{calc.}}$ from eqn. (11)	$\eta_{\text{calc.}}$ from eqn. (12)	$\eta_{\text{obs.}}$
80	85.78	16.76	0.00250	0.00253	0.00246
100	82.15	14.36	0.00214	0.00211	0.00208
120	77.53	12.04	0.00179	0.00176	0.00178
140	72.24	9.80	0.00148	0.00148	0.00153
160	65.91	7.66	0.00124	0.00125	0.00129
180	59.87	5.62	0.00108	0.00108	0.00109

### References.

- (1) Verschaffelt, "Commun. of Leiden," no. xxviii. (1896); Sugden, J. Chem. Soc. p. 1780 (1927).
- (2) Ferguson, Phil. Mag. xxxi. p. 37 (1916); Ferguson and Kennedy, Trans. Farad. Soc. xxxii. p. 1474 (1936).
- (3) Jones and Bowden, Phil Mag. xxxvi. p. 705 (1945).
- (4) Kleeman, Phil. Mag. xxi. p. 83 (1911); Macleod, Trans. Farad. Soc. xix. p. 38 (1923).
- (5) J. Phys. Chem. xxxii. p. 576 (1928).
- (6) Tyrer, J. Phys. Chem. xvii. p. 734 (1913); Saha and Srivastava, 'A Textbook of Heat,' p. 220 (Allahabad, 1931; The Indian Press).
- (7) Journ. Amer. Chem. Soc. xxxi. p. 1099 (1909).
- (8) Proc. Roy. Soc. Dublin, xii. p. 374 (1910).
- (9) Journ. Amer. Chem. Soc. lviii. p. 1139 (1936).
- (10) 'International Critical Tables' (New York, 1927), Vol. II. p. 458; Vol. III. p. 206.
- (11) *Compt. rend. Acad. Sci. U.R.S.S.* xxxi. p. 233 (1941).
- (12) J. Chem. Soc. cxxv. p. 38 (1924).

University College,  
Cardiff.

*XLVIII. The Measurement of Young's Modulus and its Temperature Coefficient for Short Filaments by Flexural Vibrations.*

By Y. L. YOUSEF, M.Sc., Fouad I. University, Cairo\*.

[Received October 11, 1945.]

ABSTRACT.

Young's modulus and its temperature coefficient between 30 and 200° C. have been determined dynamically for a number of materials in the form of thin wires (or strips). The elasticity modulus is calculated in terms of the density, the radius (or the thickness), the length and the natural frequency of transverse vibrations when one end is rigidly clamped. The density is found by weighing a specimen in air and in a suitable liquid. The average radius (or thickness) is estimated from the mass, the density, the length (and the breadth) of the test piece. The frequency to about 2 per thousand is measured for amplitudes of the order of 0.0005 cm. by vibrating the wire across a narrow beam of light falling on a photoelectric cell; the modulations in the photo-current are amplified by a 3-stage resistance-capacity coupled amplifier and are fed into one pair of plates of a cathode-ray oscillograph, whose other pair of plates is connected to a calibrated beat-frequency oscillator. A very gentle touch on the support of the wire sets up damped vibrations, which can be maintained by frequent tapping. The superposition of the two perpendicular oscillations produces Lissajous' figures, and the vibration frequency can be read directly if we adjust the oscillator until a "stationary" loop is shown on the oscillograph. The range of frequency used lies between 40-500 c.p.s. Measurements can be made on filaments as short as 0.5 cm. and as thin as 0.001 cm. On account of the very small region occupied by the test sample, it is also possible to make measurements in strong magnetic fields. Further, the same arrangement can be used for the measurement of small amplitudes of vibration. The work confirms the view that the temperature coefficient is large for materials with low melting points.

1. *Introduction.*

It can be shown <sup>(1)</sup> that for a clamped-free rod, the natural frequency  $n$  of lateral vibrations is connected to the length  $l$ , the density  $\rho$  and Young's modulus  $E$  by the relation

$$E = \frac{4\pi^2\rho}{m^4k^2} n^2 l^4, \quad . . . . . (1)$$

where  $m$  is an abstract number whose value for the gravest mode of vibration is 1.8751. For a bar with a circular cross-section of radius  $a$ ,  $k^2 = a^2/4$ . For a bar with a rectangular cross-section having a thickness  $d$  in the plane of vibration,  $k^2 = d^2/12$ .

---

\* Communicated by the Author.

It will be realized that the evaluation of  $E$  required fundamentally the measurement of  $n$ .

An early attempt in this respect was made in 1929 by Prosad <sup>(2)</sup>, who maintained the vibrations electro-magnetically and determined the frequency chronographically. Obviously, the test material must be loaded with a style, and, in case of non-magnetic substances, also with a stalloy attachment. A correction has to be made for the load. The accuracy attainable by this method is naturally not high.

In the case where the emitted note is of appreciable intensity, the frequency may be determined acoustically <sup>(3)</sup>.

Davies and James <sup>(4)</sup> determined Young's modulus for small bars by a vibrating reed method. The reed carries at the free end a stalloy strip which vibrates by the combined action of two coils wound on a soft-iron core, one carrying a polarizing D.C., and the other the driving A.C. The resonant frequency is deduced by elaborate processes, and the elasticity is calculated from a theory for loaded rods developed by Prescott <sup>(5)</sup>.

For materials of high internal friction, Grime and Eaton <sup>(6)</sup> detected the oscillations by attaching to the specimen a small piezo-electric quartz crystal. The alternating voltage set up on exciting the specimen is amplified and recorded on a cathod-ray oscillograph using a time-scale of milli-seconds.

In measuring the temperature dependence of  $E$  for  $\beta$ -brass, Bohlke and others <sup>(7)</sup> employed a reed of the material vibrating transversally by electrostatic excitation. The vibrations modulate a beam of light and thereby allow their study with a photo-electric cell and an amplifier. The natural frequency is determined by the method of resonance.

The author <sup>(8)</sup> has recently described how the frequency of a thin rod may be approximately found by measuring the time for one quarter of a vibration. In a later work <sup>(9)</sup>, a very simple method is described for the measurement of the frequency of materials which can be obtained in thin strips or fibres. The underlying principle is the formation of Lissajous' figures from the vibrating end of the fibre and a perpendicular sonometer wire of adjustable frequency, carrying a small lens opposite the fibre.

In the present work, a new method employing the principle of Lissajous' figures is used for the measurement of the frequency.

## *2. Experimental.*

The experimental arrangement is shown in fig. 1.  $L$  is a small 12-volt 0.25 ampere pilot lamp, slightly overloaded to send intense illumination. The beam is limited by a narrow rectangular slit  $S$ , 3 mm. in length and 0.25 mm. in breadth, of which a sharp image is focussed by means of a short-focus lens system  $O$  on the test material  $W$ . The latter is in the form of a wire or a ribbon, and may be kept in a temperature-controlled glass tube  $U$ . It is rigidly clamped at the upper end by means of two plane-edged aluminium jaws, which can be tightened by two nuts and screws. The clamp is firmly held by a massive stand. In case of brittle substances such as glass, or in case of soft metals such as lead, the clamp is not used, and the fixed end is cemented by litharge and glycerine into a narrow hole drilled into a lead block.





The heating coil H is wound non-inductively on the tube U, and is fixed at the ends of the tube by two copper rings. The turns are spaced by an average distance of about 3 mm. and are kept in position by a celluloid adhesive applied partially. The coil is supplied from a D.C. source.

Coupling between the amplifier circuit and any A.C. line is avoided.

Invisibly small amplitudes of vibration are used. The amplitude of the vibrating filament may be estimated by observing on the oscillograph the amplitude of the modulations, the oscillator being off. From a knowledge of the volt-sensitivity of the oscillograph (which is 7 cm./volt), and the volt-amplification of the amplifier unit, we get the amplitude of the voltage modulations at the input of the amplifier; and from a knowledge of the voltage change across the cell load (500,000 ohms) produced by known deflections of the free end of the test material, the amplitude of transverse vibrations can be readily computed. The voltage change across the load resistance produced by the displacement is evaluated from the corresponding change in the photo-electric current.

### 3. *Further Details.*

Measurements were conducted on a number of various materials to show the possibilities of the method. In particular, investigations were carried out on materials for which elasticity data are insufficient.

The substance under test is obtained in the form of a thin filament with a cross-section which is as uniform as can be secured. The density is found for a sample of the material either by weighings in air and in a suitable liquid, or by the use of a small specific gravity bottle, 5 c.c. capacity. In the cases of phosphor-bronze and platinum, only a few milligrams of the material were available, and so accurate weighings were not possible with ordinary methods. For these materials the densities were assumed from tables.

The average radius (or thickness) of the test specimen is evaluated from the mass, the density, the length (and the breadth).

A suitable length of a straight filament is clamped at one end, the section of the free end being cut at right angles to the length. The length between the free end and the flat edge of the clamp is accurately measured by a travelling microscope. The test piece is then arranged as in fig. 1 for frequency determination.

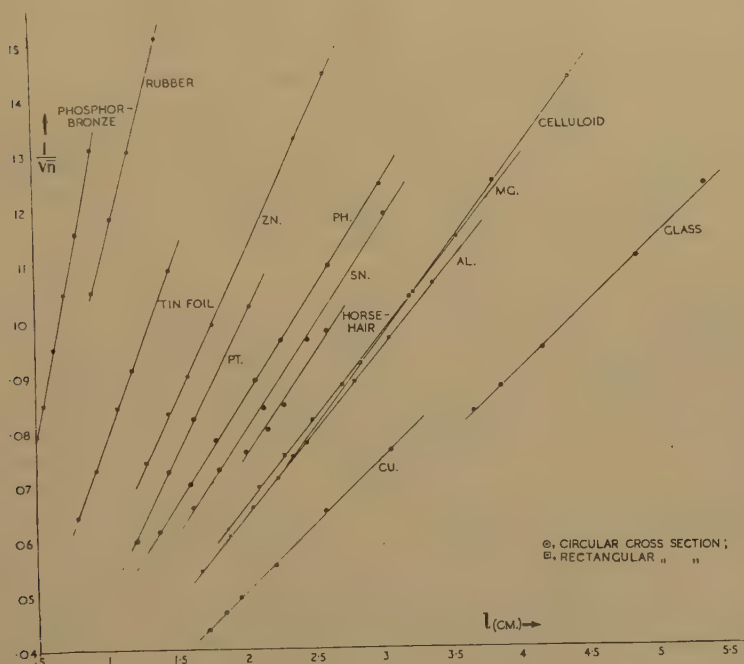
The variable audio-frequency oscillator V.A.O. is a beat frequency type (General Radio), provided with a scale reading from 0 to 21,000 c.p.s. and a vernier for fine adjustment covering the interval between  $\pm 50$  c.p.s., and readable by eye estimation to 1/10th of a cycle. There is an adjustment knob (called reed adjustment), so that when the scale is set at, say, 40 c.p.s. and the vernier at 0, the superposition of the time-controlled mains frequency (40 c.p.s.) with the oscillator frequency, when the switch K of fig. 1 is thrown to position 1, can be made to give a stationary loop on the oscillograph by a proper setting of the adjustment knob. By turning the scale knob until a stationary figure enclosing two loops is observed, we can know the scale reading corresponding to 80 c.p.s.



In measuring  $\gamma$ , we must bear in mind that partial annealing will occur at the higher temperature. Giebe and Blechschmidt<sup>(10)</sup> have shown that a change of about 18 per cent. in Young's modulus may be brought about by suitable heat treatment. It is, therefore, necessary to make the heating as small and as rapid as possible, and to check the measurement at the lower temperature. We used specimens which have been annealed for three hours at about 200° C. (in case of metals) in order to deduce the effect of further heating. The determinations of  $\gamma$  were always made at temperatures below 200° C.

In materials such as hair and celluloid, complications arise due to thermal anisotropy, which produces bending in the test piece as the temperature is raised, with the result that the effective length is reduced.

Fig. 2.



#### 4. Results.

Typical curves, connecting one observed length ( $l$ ) and the reciprocal of the square root of the frequency ( $1/\sqrt{n}$ ), are shown in fig. 2 for a number of common materials. The lengths used lie between 0.5 and 5.5 cm.; the thicknesses range between 0.001 to 0.016 cm. The natural frequencies cover the interval 40 to 500 c.p.s. The average room temperature during the measurements was 32° C.

It will be seen that, for every material, the experimental points lie well on a straight line. However, in the cases of horse-hair and glass fibre, where the uniformity of the cross-section could not be strictly ensured, there are slight discrepancies.

A summary of the results is given in Table I., where  $S$  stands for the slope of the curve, and the data are expressed in c.g.s.u. As the elasticity depends largely on the exact chemical constitution as well as on the previous physical history of the specimen, a comparison between our results and the results of previous investigators is of little value as a criterion of the accuracy of the present method.

In the present work, the frequency  $n$  could be easily measured to about 2 per thousand and the length  $l$  to better than 1 per thousand. The density  $\rho$  was determined from weighings by a chemical balance, and the accuracy depends on the mass available. In general, the accuracy in  $\rho$  is about 1 per cent. The average radius  $a$  was estimated from the mass, the length and the density, but the error in  $a$  depends primarily on the error in  $\rho$ , since the other quantities involved could be generally deter-

TABLE I.  
Young's Modulus at Room Temperature.  
(32° C.)

Material.	$\rho$ .	$a$ .	$d$ .	$S$ .	$10^{-11} E$ .
Tin .....	7.621	0.009774	—	26.44	4.98
Lead (94 per cent.) .	11.33	0.01610	—	25.51	2.363
Zinc .....	7.071	—	0.005558	18.70	10.73
Magnesium .....	1.743	—	0.01280	31.80	4.167
Aluminium .....	2.730	—	0.01245	32.53	7.561
Copper .....	8.66	0.01596	—	40.70	11.91
Phosphor-bronze ..	8.8	—	0.000995	7.96	13.68
Platinum .....	21.45	0.005015	—	19.88	17.03
Tin foil .....	7.330	—	0.004587	15.00	6.756
Glass (soda) .....	2.361	0.01241	—	42.40	6.33
Horse-hair .....	1.427	0.01172	—	26.88	0.692
Celluloid .....	1.382	—	0.395	29.71	0.2644
Rubber (red) .....	1.466	—	0.0865	10.64	0.0009629

mined to a higher degree of precision. In case of strips with rectangular cross-section, the error in the thickness  $d$  depends on the error in the breadth as well. Except in the case of phosphor-bronze suspension strip, the breadth lies around 0.1 cm., and could be measured to about 1 per cent. For phosphor-bronze, the breadth was too small, having an average value of 0.0145 cm., and the total mass available was less than 2 mgm. The calculated average thickness is 0.000995 cm., and is liable to an error of about 5 per cent., which implies a possible error of about 10 per cent. in  $E$ .

In Table II. we present a summary of typical results for the temperature coefficient of frequency and of Young's modulus. In the sixth column  $\alpha$  stands for the coefficient of linear expansion of the material and is assumed from tables. In the last column, M.P. stands for the melting point, and is also quoted from tables of physical constants to illustrate the correlation between  $\gamma$  and the melting point.



TABLE II.

The Temperature Coefficient of Frequency and of Young's Modulus.

Material.	<i>n</i> .	<i>dθ</i> ° C.	$-dn$ .	$-10^5 \frac{1}{n} \cdot \frac{dn}{d\theta}$ .	$10^5 \alpha$ .	$-10^5 \gamma$ .	M.P.° C.
Tin (pure).....	305.3	125.0	38.6	101.2	2.3	208	232.0
Lead (94 per cent.).	293.8	146.0	24.5	57.0	2.7	117	327.0
Zinc .....	142.8	141.0	8.0	40.0	2.8	83	419.4
Magnesium .....	270.4	140.0	8.3	21.9	2.6	47	651.0
Aluminium .....	263.5	155.0	8.3	20.3	2.6	43	658.0
Copper .....	133.2	154.0	3.0	14.6	1.7	31	1083.0
Platinum .....	187.0	87.0	too small	too small	0.9	too small	177.30
P-bronze .....	139.7	137.0	3.6	18.8	1.68	39	—
Glass (soda).....	129.2	192.0	too small	too small	0.85	too small	—
Horse-hair .....	173.0	92.0	13.5	170.0	—	—	—
Rubber (red) ....	41.1	84.0	5.9	171.0	—	—	—
Celluloid .....	128.6	76.5	21.6	218.0	11.0	450	—

## 5. Conclusion.

It will be noticed from Table II. that glass and platinum have a very small temperature coefficient of elasticity as well as a very small thermal coefficient of expansion. On the other hand, celluloid has a very large temperature coefficient of elasticity and a very large coefficient of thermal expansion.

Again, materials with small elasticity usually have large temperature coefficients, while metals with low melting points show appreciable change of elasticity with temperature. It appears that the correlation between  $\gamma$  and the melting point is an extension to Schaefer's<sup>(11)</sup> view that, between 15° C. and the temperature of liquid air, the order of ascending temperature coefficient for Young's modulus is also the order of descending melting point.

The author takes this opportunity to express his thanks to Dr. M. A. El-Sherbini for his valuable help during the progress of this work.

*Note added in proof.*—The values of  $\gamma$  in Table II. were calculated from eq. (2) which assumes  $dn/n$  is small. Exact calculation in the cases of tin, lead, zinc, magnesium and celluloid gives for  $-10^{-5}\gamma$  the values 191, 112, 80, 45, 410 respectively.

## References.

- (1) Lord Rayleigh, 'Theory of Sound,' i. p. 273 (1926).
- (2) Prosad, K., Phil. Mag. vii. p. 548 (1929).
- (3) Searle, G., 'Experimental Physics,' p. 60 (Cambridge, 1934).
- (4) Davies and James, Phil. Mag. xviii. p. 1023 (1934).
- (5) Prescott, 'Applied Elasticity,' p. 213 (Longmans, 1924).
- (6) Grime, G., and Eaton, J., Phil. Mag. xxiii. p. 96 (1937).
- (7) Bohlke, P., and others, Phys. Rev. lix. p. 940 (1941).
- (8) Yousef, Y. L., Proc. Phys. Soc. lvii. p. 346 (1945).
- (9) Yousef, Y. L., 'Nature,' clvii. p. 299 (1946).
- (10) Geibe, E., and Blechschmidt, E., Ann. d. Physik, ii. p. 905 (1931).
- (11) Schaefer, Deutch. Phys. Gesell. Verh. ii. p. 11 (1900).

XLIX. *A Bridge Method for the Investigation of Non-linear Resistors.*

By G. T. BAKER, M.Sc.

(Circuit Engineer, Automatic Telephone and Electric Co., Ltd.) \*.

[Received May 24, 1945.]

## SUMMARY.

The investigation of non-linearity by means of the voltage-current relationship involves many difficulties.

These are overcome by the use of two resistance parameters, the "steady" resistance ( $R_s = V/I$ ) and the "fluctuation" resistance ( $R_F = \Delta V / \Delta I$ ).  $R_s$  and  $R_F$  can be determined experimentally, and this provides a null method for evaluating the constants of non-linear resistors.

In the case of the  $V = CI^\beta$  type of resistors,  $\beta$  can be read directly from the bridge at single current values.

*Introduction.*

SEVERAL materials exhibit pure voltage non-linearity (as distinct from thermal non-linearity), but by far the most important of these are the carborundum compounds. At low voltages, these have an ohmic region, which is followed by a non-linear condition, where  $V = CI^\beta$ ;  $V$  and  $I$ , being the voltage and current and  $C, \beta$ , two constants. At higher voltages, the conduction again becomes ohmic †.

One such compound, marketed under various trade names, such as "Atmite" ‡, "Thyrite" §, "Metrosil" ||, follows the equation  $V = CI^\beta$  over a very wide current range.

*Voltage/Current Method of Investigation.*

The usual method of investigating non-linearity is by taking current readings at a series of voltages. The result is expressed as a voltage/current graph, from which the corresponding constants can be derived. In the case of the  $V = CI^\beta$  type of resistor, the results are plotted on logarithmic paper, and  $C$  and  $\beta$  obtained from the intercept on the Y-axis and the slope respectively. If the current range involved is sufficient, all materials depart from this simple relationship, and it then becomes extremely difficult to deduce a suitable equation from the voltage/current plot.

---

\* Communicated by the Author.

† Fairweather, J. I. E. E. vol. lxxxix. part 1, no. 23.

‡ Saville, Strowger J. vol. v. no. 1.

§ Brownlee, Gen. Elect. Rev. vol. xxxvii. p. 175.

|| Radley *et al*, P. O. E. E. J, vol. xxxiv. part 4,

The method is entirely dependent on the accuracy of the instruments employed, and at both ends of the current range, extremely specialized technique is necessary. To obtain close readings the current must pass for some time, and there is thus a danger of errors due to heating effects and temperature variations.

In general, the method is quite laborious and unsuitable for tests involving large numbers. Moreover, to obtain an accurate figure for  $\beta$  or an equivalent current index, it is necessary to take widely separated current readings. In this case, any variation in  $\beta$  between these values is completely masked.

#### *Resistance Parameters.*

In addition to the obvious variation of resistance with voltage, it should be appreciated that an ambiguity exists when specifying the resistance, even at a given voltage. To an engineer the resistance would be the ratio of the voltage to the current, *i. e.*  $V/I$ , which will be termed the "steady" resistance or  $R_s$ .

A mathematician, however, would probably declare the resistance at a given voltage, to be the slope of the voltage-current curve at this voltage, *i. e.*  $\Delta V/\Delta I$ , which will be termed the "fluctuation" resistance or  $R_F$  \*.

These duplex values, however, are extremely useful, since they provide a pair of parameters,  $R_s$ ,  $R_F$ , which can be used for investigating non-linearity in a manner which avoids the difficulties of voltage-current measurements.

The relationship between  $R_s$  and  $R_F$  depends entirely on the voltage-current relationship, but the equation is invariably much simpler, and quite readily obtained from an experimental plot.

As an example, consider the voltage/current equation

$$\begin{aligned} V &= CI^\beta \\ V + \Delta V &= C(I + \Delta I)^\beta \\ I + \Delta I &= (V/C)^{1/\beta} (1 + \Delta V/V)^{1/\beta} \\ &= I(1 + 1/\beta \cdot \Delta V/V \dots) \end{aligned}$$

If  $\Delta V/V$  is very small, further terms containing higher powers of this factor can be ignored.

$$\begin{aligned} \therefore \Delta V/\Delta I &= \beta \cdot V/I \\ R_F &= \beta R_s \end{aligned}$$

This linear equation is obviously very much simpler than the corresponding voltage-current equation from which it has been derived. In the case of the more general equation

$$V = RI + CI^\beta,$$

---

\* Differentials have been used throughout in place of the calculus, as being equivalent to the finite voltages used in practice, and also to show errors introduced by approximations.

which is extremely difficult to handle and to evaluate from an experimental curve, the corresponding resistance equation is the simple linear

$$R_F = R + \beta \cdot R_S.$$

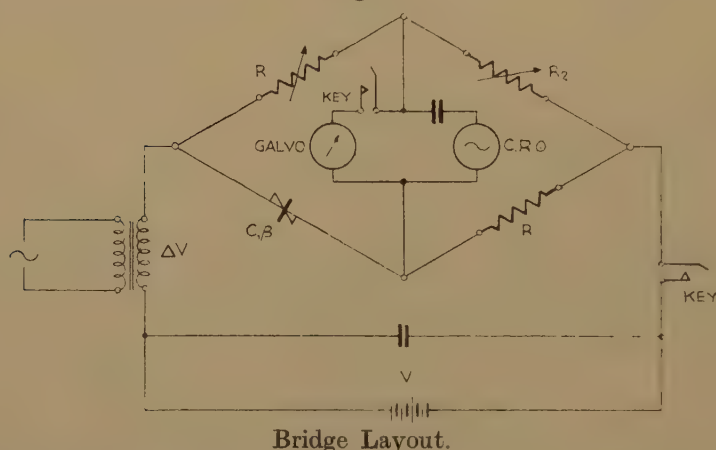
It will be found that, in every case, the resistance equation is much simpler than its current/voltage equivalent.

### *Experimental Method.*

A straightforward resistance bridge is used, with the layout shown in fig. 1. A direct voltage  $V$  is applied to the bridge and a D.C. balance gives  $R_S$ . Superimposed upon  $V$  is a low amplitude sine wave, which constitutes  $\Delta V$  and an A.C. balance on the bridge gives  $\Delta V/\Delta I$ , *i. e.*  $R_F$ .

It is obviously necessary to ensure that  $\Delta V$  is small compared with  $V$ , and that the action of balancing the bridge does not vary the steady voltage across the resistor. For this reason the bridge should always be adjusted by varying  $R_1$  or  $R_2$  (fig. 1). After the D.C. balance is obtained, the galvanometer must be disconnected before adjusting for

Fig. 1.



an A.C. balance, otherwise the current through the galvanometer disturbs the potential across the resistor. A condenser in the A.C. balancing circuit prevents D.C. flowing through this path.

The D.C. balance is obtained by means of a centre-reading galvanometer, and a cathode-ray oscillograph or an amplifier with headphones gives the A.C. balance. The former is preferable for reasons discussed later.

Other details of the bridge depend largely on its specific purpose. The typical apparatus described in detail below was intended for rapid testing of commercial resistors of the  $V=CI^\beta$  type.  $\beta$  was the important constant and, in addition, it was desired to know the current passed at a given working voltage. This can be obtained from  $R_S$  and accordingly the bridge was designed to give this value, also  $\beta$ , *i. e.*  $R_F/R_S$ . Both can be read directly from the bridge without calculation or plotting.



*A Typical Bridge.*

Referring again to fig. 1,  $R_1$  and  $R_2$  are two variable resistors. The ordinary, wire-wound, circular-slide type is quite suitable since a simple calibration method is given below. A good quality, non-inductive model, with a low temperature coefficient is necessary for accurate work.  $R_1$  should be about 100,000 and  $R_2$  approximately 6,000 ohms, the actual values being unimportant, since all measurements are on a ratio basis.  $R$  is a fixed resistance approximately 1000 ohms.

The bridge is calibrated as follows:—The non-linear resistor marked  $C, \beta$ , is replaced by a graduated resistance box of the decade type.  $R_1$  is set to a maximum, and the resistance box adjusted to the nominal maximum of  $R_1$ , say 100,000 ohms. A D.C. balance is obtained on the bridge by adjusting  $R_2$ . The position of its pointer at balance is recorded and marked "1." The resistance box is reduced to 90,000 and the bridge again balanced by altering  $R_2$ . This point is marked ".9." In this manner, the resistance box is reduced in stages. At each stage a balance is obtained, and the ratio marked on the scale of  $R_2$ . The completed scale should read from about "0.17" to unity.

The pointer of  $R_2$  is returned to the position marked "1," and the resistance box to 100,000 ohms. The bridge should balance and the pointer of  $R_1$  is marked 100,000. The box is again reduced in stages, but this time  $R_2$  remains at unity and a balance is obtained on  $R_1$ . At each point of balance the resistance is marked on  $R_1$ , which is thus scaled from "0" to "100,000" ohms.

The direct voltage used on this bridge varied between 30 and 70 volts, and the alternating voltage was  $1/4$  V. at 1000 C.P.S. The latter frequency was chosen as being sufficiently high to ensure simple amplification, without involving capacity and leakage difficulties. It is advisable to test the first few samples over a range of frequencies; any variation in the results obtained indicating that the resistances are not of a completely non-inductive nature.

The method of operation is as follows:—The resistor is connected to the appropriate terminals of the bridge and  $R_2$  set at unity. The galvanometer key is thrown and the direct voltage applied.  $R_1$  is adjusted to give a D.C. balance and this reading gives  $R_s$ . The alternating voltage should be applied throughout. The galvanometer key is opened, and with  $R_1$  still in the balance position an A.C. balance is obtained by varying  $R_2$ . The reading on this scale at the point of balance gives  $\beta$  directly.

This bridge has been described in some detail, merely to illustrate the fact that by manipulation of the various components the bridge can be made to give the basic information required in a very direct form. For general investigation the associated values of  $R_s$  and  $R_F$  are probably more useful. In this case a straightforward Wheatstone bridge is used, and the balance obtained by adjusting  $R_1$  in each case.  $R_2$  would then be used to vary the ratio in the well-known manner.

*Accuracy.*

The resistances comprising the bridge should, of course, be accurate, non-inductive, and have a low temperature co-efficient. With these basic points satisfactory, the accuracy that can be obtained depends largely on the amount of amplification that can be introduced into the A.C. balancing circuit. With a high gain at this point the amplitude of the input,  $\Delta V$ , can be low, and a fine degree of balance obtained.

For this reason, all wiring should be short and run in balanced pairs with earthed screening. The most dangerous wires are the long battery leads, and a condenser of several micro-farads should be connected across these at the bridge to avoid pick-up. With these precautions no difficulty should be countered when using sufficient amplification to obtain  $\beta$  to three significant figures.

The cathode-ray oscillograph was found to be the best balancing medium, possibly because these experiments were conducted in a noisy laboratory, but also because it turns to good account a factor which becomes a defect in the audible method of balancing.

The first ignored term in the expansion used as the basis of this method is

$$\frac{1}{2} \cdot 1/\beta \cdot (1/\beta - 1)(\Delta V/V)^2 \sin^2 \theta.$$

Now  $\sin^2 \theta$  expands into

$$\frac{1}{2} + \frac{1}{2} \cos 2\theta.$$

The first item indicates an apparent increase in the direct voltage and does not affect the accuracy, provided the A.C. is applied throughout as advised earlier.

The other term is a second harmonic which is not balanced in the linear arm of the bridge. Although its amplitude is small, the amplification at the balance point can be increased until it is visible on the trace. Under these circumstances, the wave-shape (rendered stationary by the oscillograph time base) undergoes quite a sudden change as the balance point is reached, since the associated fundamental moves through  $180^\circ$ . This enables a very fine degree of balance to be obtained and renders the method independant of the second term of the expansion.

L. *Dr. Löwy's Theory of Ground-water Accumulation.*

*To the Editors of the Philosophical Magazine.*

GENTLEMEN,—

Dr. LÖWY's theory of ground-water accumulation (Phil. Mag. xxxvi. p. 651) is vitiated by an erroneous assumption of an infiltration velocity for the downward movement of water in soil. The error arises because the soil interstices are pictured as a system of tubes instead of a system

of cellular pore spaces connected with one another by narrow necks. The consequences of this pore-space geometry have been worked out both theoretically and experimentally at Rothamsted by Haines<sup>(1)</sup> and by Schofield<sup>(2)</sup>, who have shown that water in soil is subject to hysteresis. Haines' treatment was based on the surface tension and pressure deficiency under the water menisci and was, therefore, strictly applicable only to coarse-grained material in which confirmatory experiments were feasible, but Schofield's treatment being based on energy relationships was general, and therefore applicable to all soils from heavy clays to sands. One of the consequences of the hysteresis is, of course, that the water tends to resist the movements caused by forces such as evaporation and percolation. The picture can be given simply as follows:—

Consider a small quantity of rain falling on the surface of a dry soil. The thin top layer will absorb a certain amount. The excess passes to the next layer until its absorptive capacity is satisfied, and so on until all the rain has entered the soil. At this stage further downward movement of the water front virtually ceases because it would involve a thinning of the water films, and thus an increased pressure deficiency resisting the movement. The water content in these conditions is known as the field capacity. The value is not unique for a given soil in a given state of packing, for it depends obviously on such factors as the actual amount of rain, and whether the soil was initially absolutely dry, and it has a smaller value than the value for complete saturation of the pores. Movement of water, either by distillation or by liquid creep, from the zone of field capacity into the adjoining soil does not entirely cease, but it is very slow and extends only to small distances. Experiments by Veihmeyer<sup>(3)</sup> and Bodman and Colman<sup>(4)</sup> on the wetted zone of surface irrigated soils provide good experimental proof of this.

In general, therefore, a small quantity of rain, once it has entered a dry soil, is necessarily held in and just below the soil surface. It will not, however, evaporate as quickly under the intense desert conditions as might be supposed at first sight. The curvature of the films in the surface will increase by evaporation, and water will begin to move towards them from the adjacent thicker films. Once the loss by evaporation exceeds the rate at which replacement water arrives, the top surface becomes dry. Thereafter the water evaporates *in situ* and the soil dries out from above downwards. The low heat conductivity of this increasingly thick dry surface layer protects the moist layer below from the full force of evaporation. For the reasons given above the transition between the dry layer and the moist soil below is quite sharp.

If large underground water deposits exist in desert conditions, that are not due to subterranean flow from distant catchment areas, they are most likely due to local variations in topography. A considerable proportion of a small but heavy rainfall will not be directly absorbed but run over the surface to a lower level. A slight depression will accumulate and absorb a quantity of run-off water corresponding to many times the actual rainfall,



A fuller account of the points dealt with above will be found in my Presidential Address <sup>(5)</sup> to the Royal Meteorological Society for 1939.

Yours faithfully,

B. A. KEEN.

### *References.*

- (1) J. Agric. Sci. xx, p. 97 (1930).
- (2) Trans. 3rd Int. Cong. Soil Sci. ii. p. 37 (1935).
- (3) Trans. Amer. Geoph. Union, pt. ii. p. 613 (1938).
- (4) Proc. Soil Sci. Amer. viii. p. 116 (1944).
- (5) Q. J. Roy. Met. Soc. lxv. p. 123 (1939).

Rothamsted Experimental Station,  
Harpenden, Herts.

### *LI. Notice respecting New Book.*

*The Diffraction of X-rays and Electrons by Free Molecules.* By M. H. PIRENNE.  
[Pp. xii+160, with 2 plates.] (Cambridge University Press, 1946.) Price 12s. 6d.

THE pioneer work of von Laue and the Braggs, by opening up the new fields of structural and chemical crystallography, made it possible for the exact arrangement of atoms and molecules in the solid state to be determined, and shed valuable light on the nature of chemical bonds. But the structural study of free molecules in the gaseous state has proved a much more difficult matter. Even now, after over 30 years of theoretical and experimental investigation, the number of molecules studied can be listed in one short table.

Dr. Pirenne has performed a service to both physicists and chemists by giving a logical account both of the theoretical basis of the study of X-ray diffraction by gases (with enough but not too much mathematical detail) and of the information it has given about the structure of atoms and molecules. Experiments on the scattering power of helium, for example, have shown very clearly the inadequacy of the old Bohr orbital model of the atom, and the superiority of the new wave-mechanical treatment; for atoms do not scatter X-rays as if they were points, and the atomic scattering curve is dependent on the electron density distribution. The diffraction of X-rays by free molecules is also dependent upon, and will therefore give information about, the electron density distribution in the molecule, which in turn depends on the relative positions of the atoms. The use of fast electrons instead of X-rays gives, or can give, information even about the positions of H atoms, in molecules which also contain heavy atoms.

Dr. Pirenne has very properly related his subject to the problems of diffraction in crystals and liquids. He gives an adequate bibliography and the book is well indexed and produced with the care and style always associated with the name of the Cambridge University Press.

K. L.

[The Editors do not hold themselves responsible for the views  
expressed by their correspondents.]

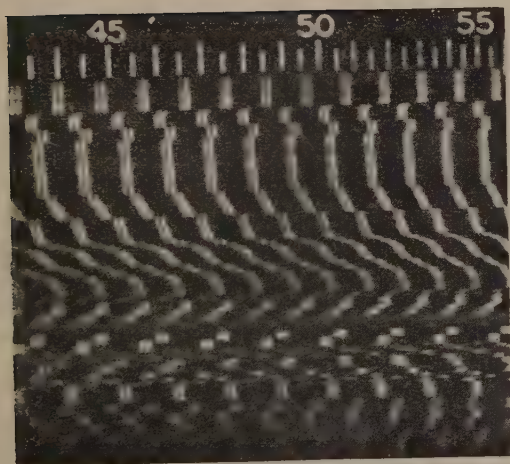




(a)



(b)



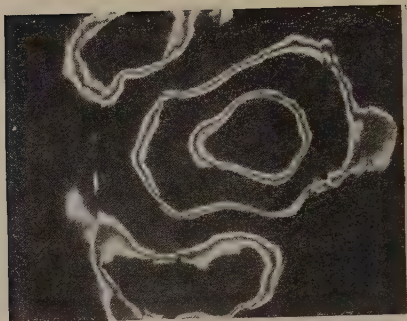
(c)



(d)



(a)

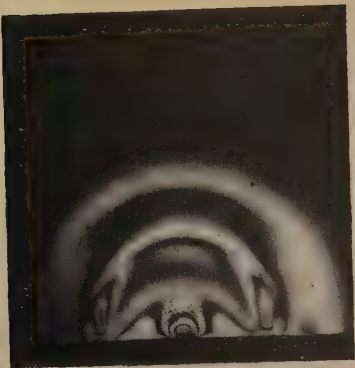


(b)

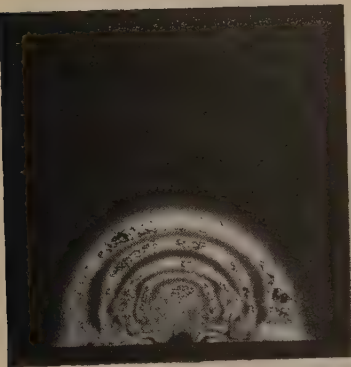


(c)





1.  
0 Microseconds.



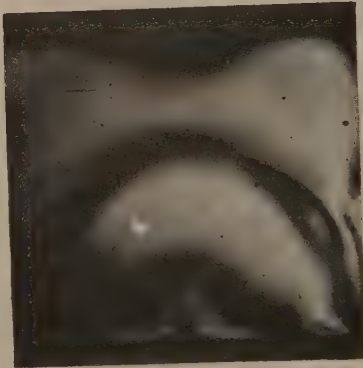
2.  
16 Microseconds.



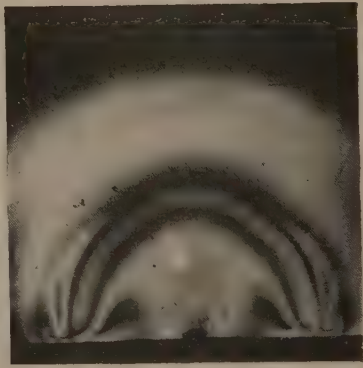
3.  
33 Microseconds.



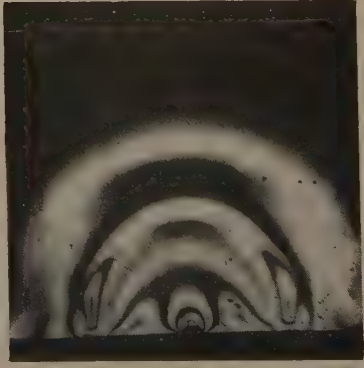
4.  
43 Microseconds.



5.  
80 Microseconds.



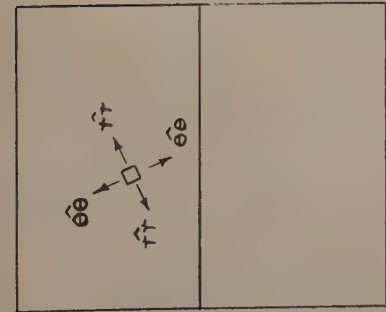
6.  
70 Microseconds.



7.  
53 Microseconds.

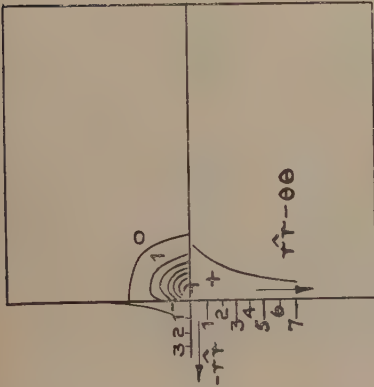
Pellet Striking Edge of Wide Plate.





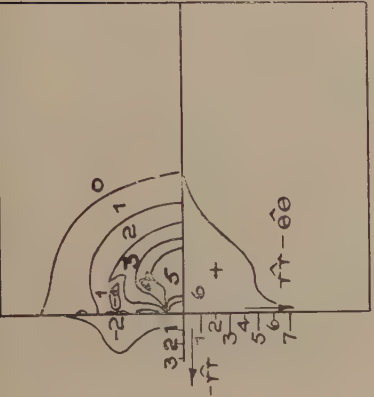
1.

0 Microseconds.



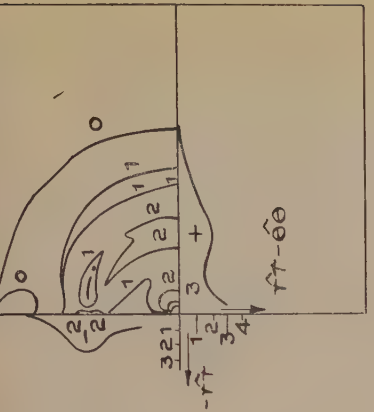
2.

16 Microseconds.



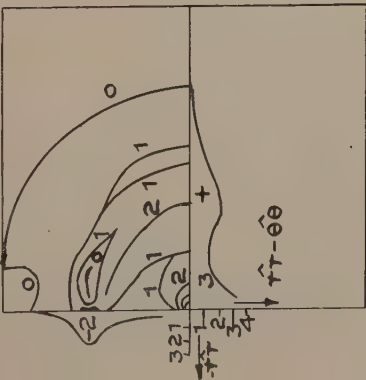
3.

33 Microseconds.



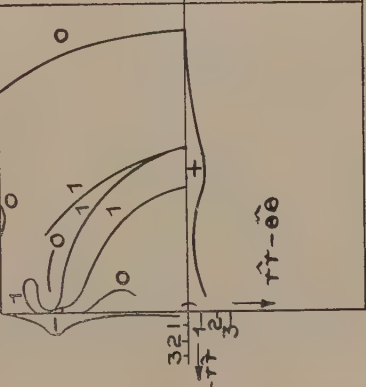
4.

43 Microseconds.



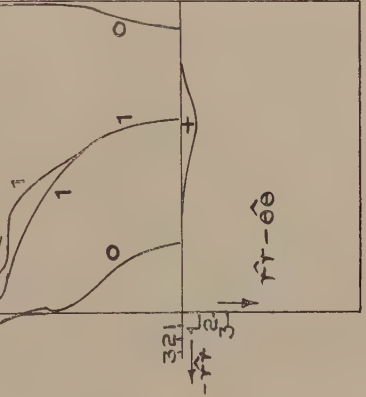
5.

53 Microseconds.



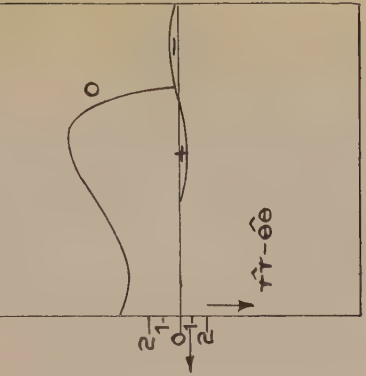
6.

70 Microseconds.



7.

80 Microseconds.



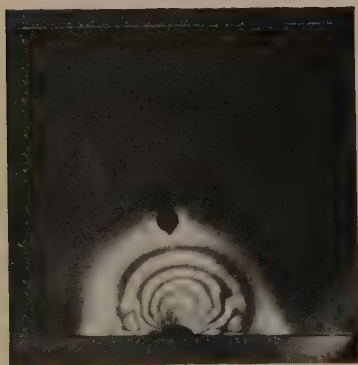
8.

90 Microseconds.

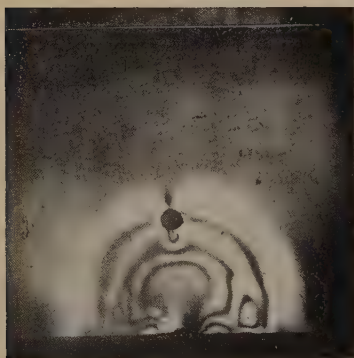
Principal Stress Differences corresponding to Fringe Photographs in Plate IX. Stresses plotted as Fringe Orders.



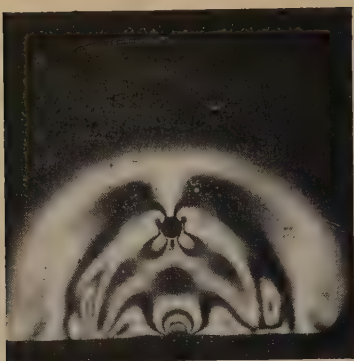
1.  
16 Microseconds.



2.  
30 Microseconds.



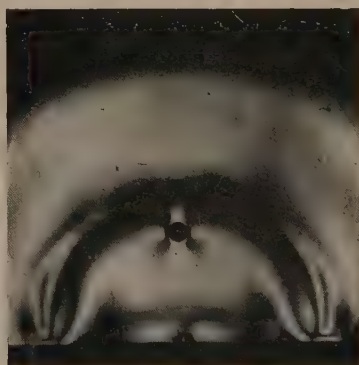
3.  
40 Microseconds.



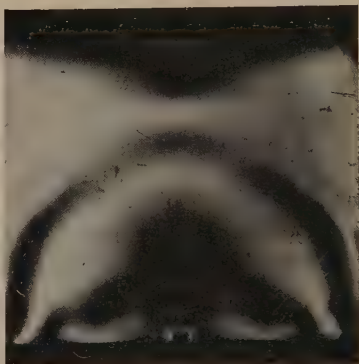
4.  
48 Microseconds.



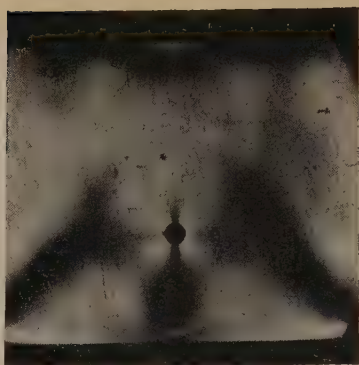
5.  
51 Microseconds.



6.  
68 Microseconds.

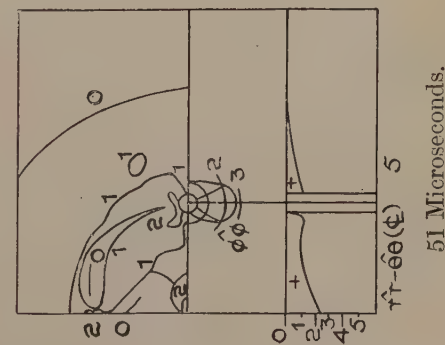
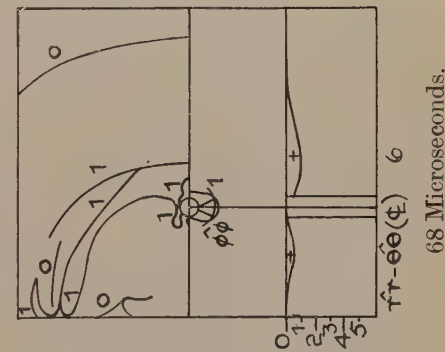
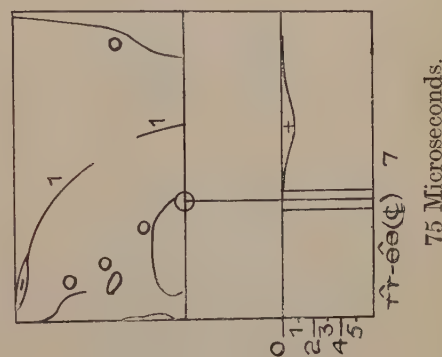
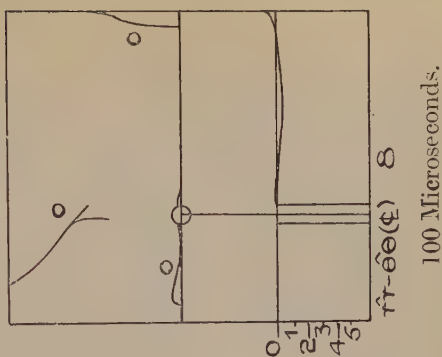
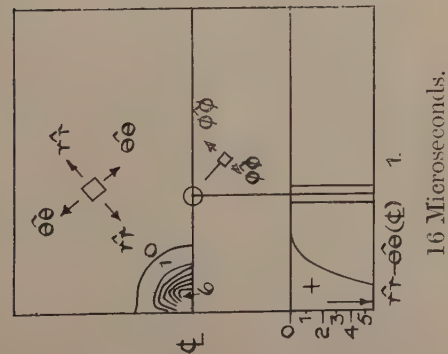
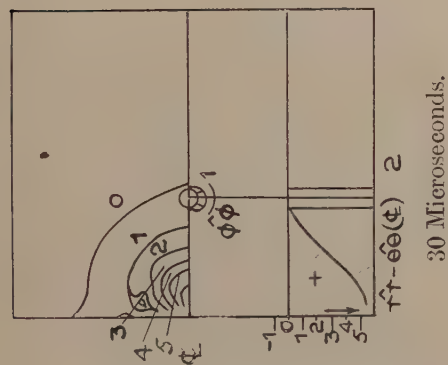
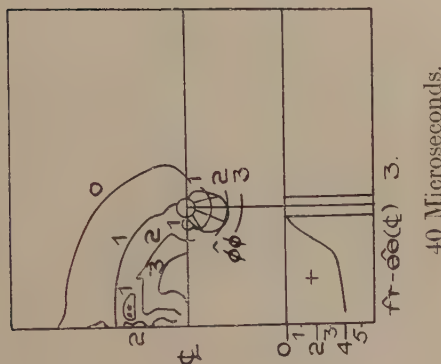
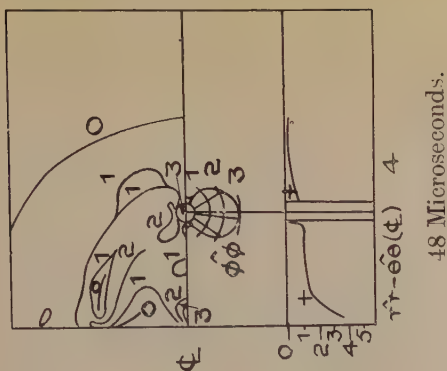


7.  
75 Microseconds.



8.  
100 Microseconds.

Stress Wave passing Small Hole.



Principal Stress Differences corresponding to Fringe Photographs in Plate X. Stresses plotted as Fringe Orders.



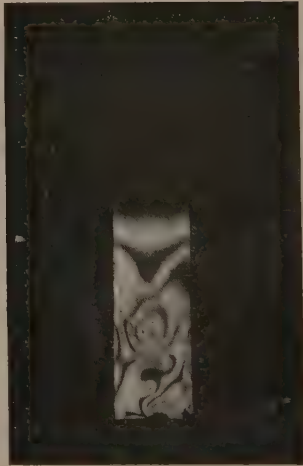
1.  
12 Microseconds.



2.  
30 Microseconds.



3.  
60 Microseconds.

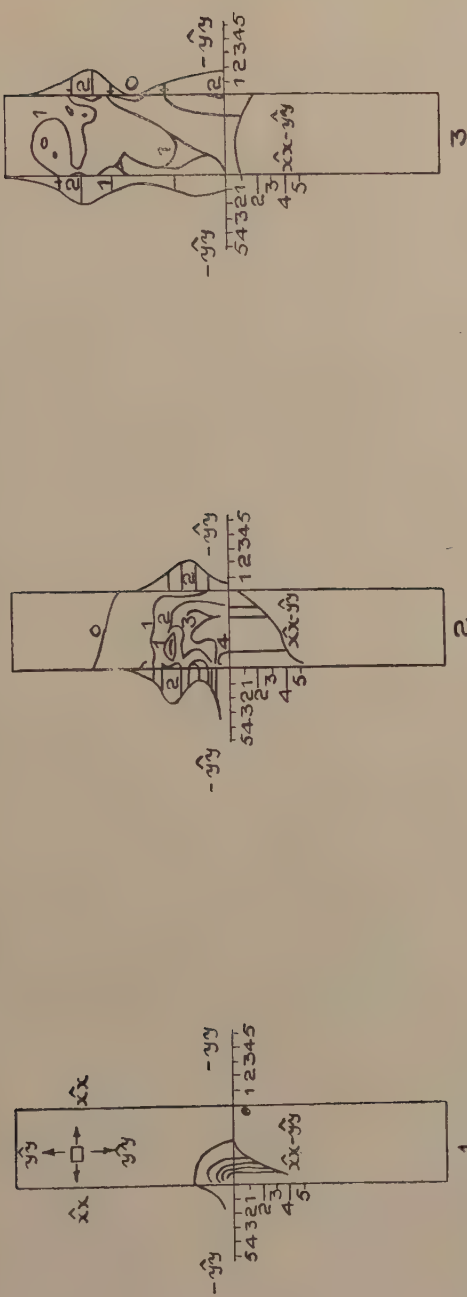


4.  
60 Microseconds.

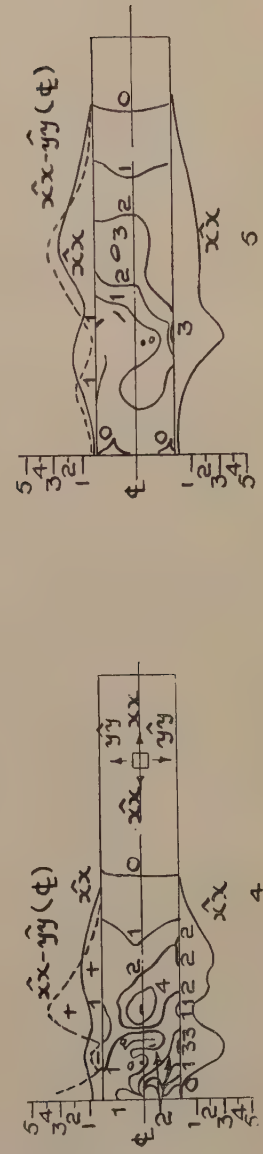


5.  
90 Microseconds.  
Transverse and Longitudinal Impact of Pellet with Rectangular Section Bar.



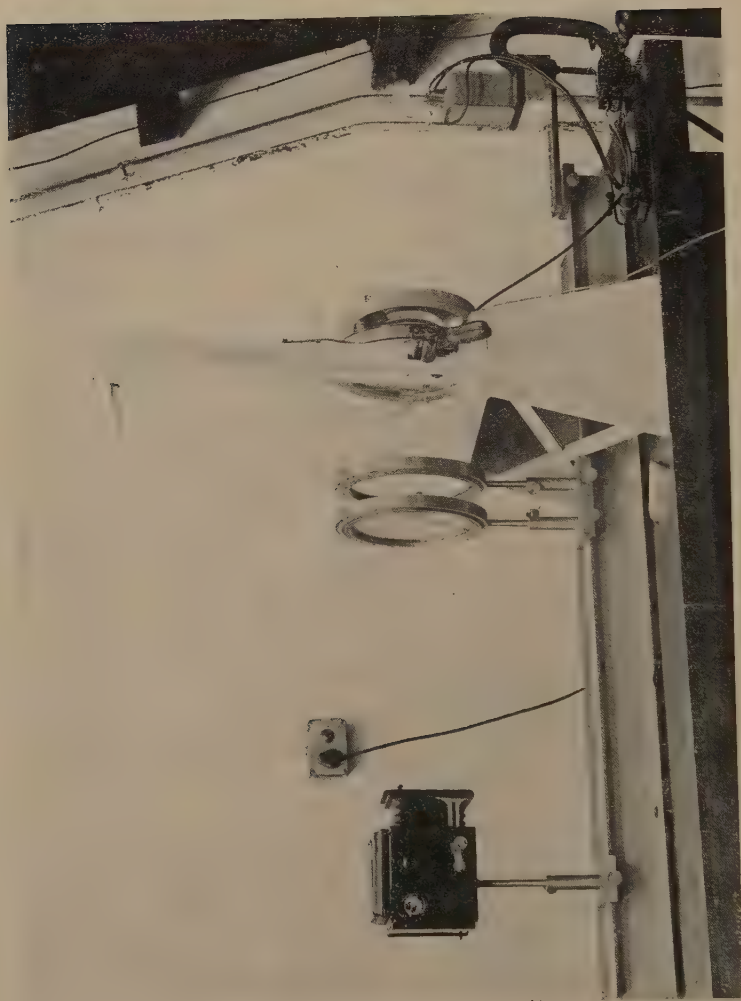


12 Microseconds. 30 Microseconds. 60 Microseconds.



60 Microseconds. 90 Microseconds.

Principal Stress Differences corresponding to Fringe Photographs of Plate XI. Stresses plotted as Fringe Orders.



General view of Stress Wave Apparatus.

LII. *Further Notes and Suggestions on the Teaching of Physics.*

By C. J. SMITH, Ph.D., M.Sc., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey \*.

[Received May 31, 1945.]

IN the following paper a detailed discussion is given of an important graphical method for determining the constants which appear in certain types of equation often met with in laboratory practice. Other graphical methods are also indicated. In addition, a more complete theory of Sents' method of determining the surface tension of a liquid is given. The theory of the steady flow of a liquid through a vertical capillary tube is worked out from first principles and then applied to a new design of such a viscometer for students' use, and also to two forms of Ostwald viscometer. The fundamental properties of the focal planes of a coaxial system of refracting surfaces are derived from elementary principles. Finally, an attempt is made to elucidate the fundamental principles underlying the construction and use of some electrical instruments. Several other relevant topics are also raised and discussed.

## PART I.

*On the curve  $y-b=\kappa(x-a)^n$ .*

In his textbook 'Elementary Practical Mathematics,' Prof. John Perry has given a beautiful graphical method of constructing the curve represented by the above equation when the position of one point on it is known. Unfortunately, Perry has made a slip which leads to much confusion: the slope of a line is taken to be synonymous with the tangent of an angle and, in addition, this author has discussed the problem from the following points of view: (a)  $n$  is positive, (b)  $n$  is negative. This latter restriction is not necessary if proper attention is paid to signs, so that the argument may be amended as follows:—

Let  $Ox$ ,  $Oy$ , fig. 1, be the usual rectangular axes and let  $P_1$  be a point on the curve  $y-b=\kappa(x-a)^n$ . Let  $OA=a$ ,  $OB=b$ , and through  $A$  and  $B$  draw the straight lines  $AH$  and  $BK$  with slopes  $\alpha$  and  $\beta$  respectively, the co-ordinates axes being considered as lines of zero slope. Let  $M$  be the projection of  $P_1$  on  $Ox$ , and through  $M$  draw  $ML$  with slope  $\theta$  to cut  $AH$  in  $L$ . Let  $N$  be the projection of  $L$  on  $Ox$ . Similarly, let  $R$  be the projection of  $P_1$  on  $Oy$ : then draw  $RQ$  with slope  $\phi$  with respect to  $Oy$ ,  $\beta$  and  $\phi$  being considered positive when the straight lines  $BK$  and  $RQ$  are as indicated. If  $S$  is the projection of  $Q$  on  $Oy$ , then  $NL$  and

---

\* Communicated by the Author.

As illustrations, the following examples are cited:—

(i) The frequency,  $f$ , to which a Helmholtz resonator of volume  $v$  responds, is known to be represented by the equation  $f = \kappa v^{-\frac{1}{4}}$ , where  $\kappa$  is a constant. Now in practice the "volume"  $v$  is uncertain to an amount  $\epsilon$ , say, so that we may write  $f = \kappa(v + \epsilon)^n$ , where  $\kappa$ ,  $\epsilon$ , and  $n$  are three constants to be determined. In ordinary laboratory practice it is usual to assume that  $\epsilon \rightarrow 0$  and so discover  $n$ , or else assume  $n$  to be  $-0.5$  and so find  $\epsilon$ . Now the first assumption is untrue and the second is not necessary, as is revealed by the following analysis of some observations made with the apparatus shown in fig. 3 (a). The resonator consisted of a cylindrical brass box of the dimensions indicated. The fork of highest frequency [512 cycle. sec.<sup>-1</sup>] was struck and then held at least 6 cm. away from the opening of the resonator. This was initially filled completely with water which was run out into a measuring cylinder until resonance occurred. If the mark was "overshot," the water could easily be returned to the resonator: a similar process enabled observations to be repeated. The glass tube attached to the resonator was given the form shown so that it could be rotated about a horizontal axis and air bubbles easily removed. The following is a set of observations obtained with this apparatus: Let  $V$  denote the volume of the resonator measured from its open end. Then we may write  $f = \kappa(V - a)^n$ , where  $a$  is a correction term.

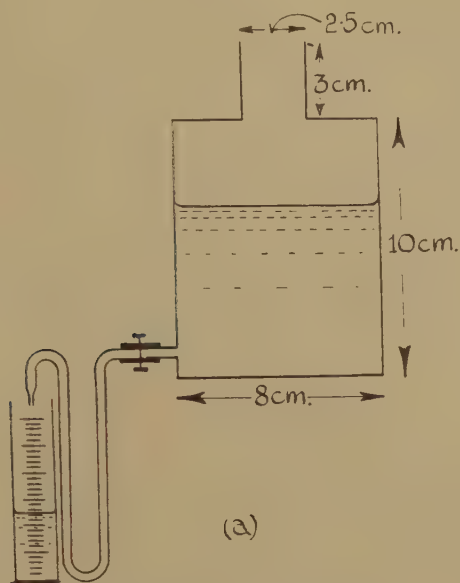
Frequency of fork [ $f$ cycle. sec. <sup>-1</sup> ]	Volume of water [ $V$ cm. <sup>3</sup> ]	Frequency of fork [ $f$ cycle. sec. <sup>-1</sup> ]	Volume of water [ $V$ cm. <sup>3</sup> ]
512	$\left. \begin{array}{l} 123 \\ 123 \\ 124 \end{array} \right\} 123.3$	341.3	$\left. \begin{array}{l} 263 \\ 264 \\ 263 \end{array} \right\} 263.0$
480	$\left. \begin{array}{l} 136 \\ 138 \\ 137 \end{array} \right\} 137.0$	320	$\left. \begin{array}{l} 294 \\ 296 \end{array} \right\} 295.0$
426.6	$\left. \begin{array}{l} 172 \\ 172 \\ 172 \end{array} \right\} 172.0$	288	$\left. \begin{array}{l} 358 \\ 360 \\ 358 \\ 357 \end{array} \right\} 358.3$
384	$\left. \begin{array}{l} 206 \\ 207 \\ 208 \\ 208 \end{array} \right\} 207.5$		

The observations were plotted on a large sheet of paper and a smooth curve drawn through them with the aid of a spline: it is reproduced as CD in fig. 3 (b). The line of slope  $\beta = \frac{1}{3}$  was drawn by plotting the points  $Q' = (-100, 300)$ ,  $Q'' = (-200, 600)$ , and a preliminary drawing, not reproduced here, indicated that the lines  $R_1Q_1$ , etc., and  $M_1L_1$ , etc., could be drawn with the aid of a  $45^\circ$  and a  $60^\circ$  set-square respectively. Beginning with a convenient point  $P_1$ ,  $P_1R_1$  was drawn parallel to  $Ox$ ,  $R_1Q_1$  at an actual angle of  $135^\circ$  to  $Oy$ , and  $Q_1P_1'$  parallel to  $Ox$ .  $P_1M_1$



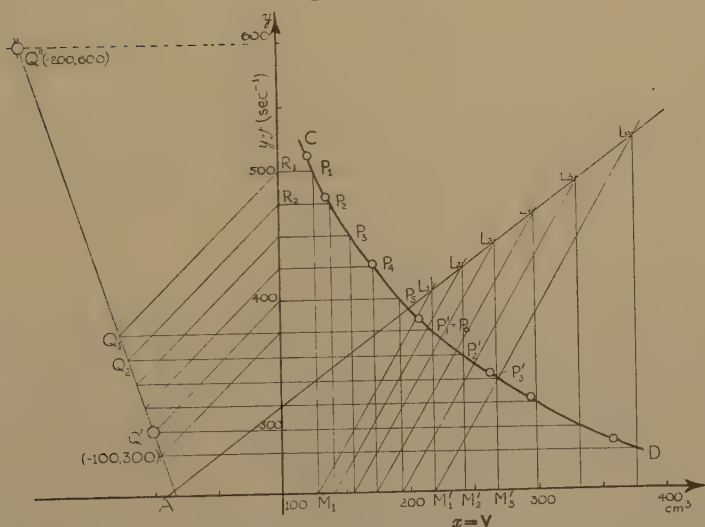
and  $P_1'M_1'$  were drawn perpendicular to  $Ox$ ; and  $M_1L_1$  drawn with the aid of a  $60^\circ$  set-square. Thus  $L_1$  was found.

Fig. 3 (a).



A Helmholtz resonator.

Fig. 3 (b).



[N.B.—The portion of the figure between  $x=0$  and  $x=100$  is omitted to save space.]

To find  $L_2$  a second point  $P_2$  was chosen and the above construction repeated. Similarly for  $L_3, L_4$ , etc. The best straight line was then drawn through the  $L$ 's. Its intercept on  $Ox$  gave  $a=10 \text{ cm}^3$ . Also

$\alpha=0.750$ . Since  $\theta=\sqrt{3}$ ,  $\beta=\frac{1}{3}$ , and  $\phi=-1$ , we have from the equation  $[1-(\alpha/\theta)]^n=[1-(\beta/\phi)]$ ,  $n=-0.506$ .

(ii) The equivalent conductivity,  $\lambda$ , of an aqueous solution of an electrolyte is known to be related, at least approximately, to the equivalent conductivity at infinite dilution,  $\lambda_0$ , and the concentration  $c$  of the dissolved substance, by the equation  $\lambda=\lambda_0-Bc^n$ , where  $B$  and  $n$  are constants. One method of determining these constants has been given by Ferguson and Vogel \*; the constants may also be found and, I would suggest, much more readily, by the method now advocated, because the above equation is of the type  $y-b=\kappa x^n$ , where  $b=\lambda_0$ ,  $y=\lambda$ ,  $\kappa=-B$ , and  $x=c$ .

The method has been applied to the observations made by Shedlovsky † on the equivalent conductivity of potassium chloride in aqueous solution. Using the following values, as given by Shedlovsky,

$c \times 10^4$ [gm.-equiv. litre <sup>-1</sup> ]	$\lambda$ [ohm <sup>-1</sup> cm. <sup>-1</sup> gm.-equiv. <sup>-1</sup> litre]	$c \times 10^4$	$\lambda$
0.32576	149.33	11.321	146.76
1.0445	148.91	14.080	146.46
2.6570	.38	15.959	.26
3.2277	.19	20.291	145.72
3.5217	.12	20.568	.71
4.6948	147.89	23.379	145.48
6.0895	.52		
8.4200	.23		
9.2856	147.07		

it is found that  $\lambda_0=149.9$  and  $n=0.488$ . Shedlovsky gives  $\lambda_0=149.82$ .

Here it must be pointed out that the author of this present paper makes no claim that the values of  $\lambda_0$  and  $n$  he has obtained are better than those given by Shedlovsky, for, in the first place, the Debye-Huckel-Onsager theory suggests that at infinite dilution  $n$  must be 0.5. It is now only claimed that the equation,

$$\lambda=149.9-86c^{0.488},$$

does fit the curve, but it is realized that this formula has little or no theoretical value: the index 0.488 and perhaps the constant 86 must only be regarded as mean values which provide an equation to fit the curve obtained experimentally. The author hopes to discuss elsewhere, at a not-too-far distant date, another equation which is an extension of that provided by the modern theory of infinitely dilute solutions.

(iii) In the method for determining the specific heat of a liquid at a definite temperature, as developed by Ferguson ‡ and his co-workers, it

\* Ferguson and Vogel, *Phil. Mag.* l. p. 971 (1925).

† Shedlovsky, *J. A. C. S.* vol. liv. p. 1411 (1932).

‡ Ferguson and Miller, *Proc. Phys. Soc.* vol. xlv. p. 194 (1933); Cockett and Ferguson, *Phil. Mag.* vol. xxix. p. 185 (1940).

is assumed that the rate of fall in temperature of a calorimeter and its contents, when they are allowed to cool in an enclosure from an initial temperature  $\theta_0$ , obeys the law

$$-\frac{d\theta}{dt} = \kappa\theta^n,$$

where  $\kappa$  and  $n$  are constants. Integrating we get  $\theta_0^m - \theta^m = m\kappa t$ , where  $m = -n + 1$ . This equation may be written

$$t = (\theta_0^m / m\kappa) - (\theta^m / m\kappa),$$

or

$$y = A - Bx^m,$$

where  $A$  and  $B$  are constants and  $x = \theta$ , which is the type now discussed.

(iv) The variation of the surface tension,  $\gamma$ , of a liquid with temperature  $\theta$ , as measured on the centigrade nitrogen scale, is related to its surface tension  $\gamma_0$  at  $0^\circ \text{C}$ . by the equation  $\gamma = \gamma_0(1 - c\theta)^n$ , where  $c$  and  $n$  are constants.

Calling  $x = \gamma/\gamma_0$  and  $y = \theta$ , the above equation may be written

$$y = (1/c) - (1/c)x^{1/n}.$$

This is an equation of the type now under discussion so that the constants  $c$  and  $n$  may be found. The method is easier than that used by Ferguson \* and, in addition, avoids the construction of tangents to a curve and the assumption that  $c = \theta_c^{-1}$ , where  $\theta_c$  is the critical temperature for the liquid concerned,

(v) Slotte's equation representing how the viscosity  $\eta$  of a liquid varies with the temperature  $\theta$  ( $^\circ\text{C}$ .) is  $\eta = \eta_0(1 + b\theta)^n$ , where  $b$  and  $n$  are constants. This equation may be written

$$[\eta/\eta_0] = b^n[\theta + (1/b)]^n,$$

so that if we write  $y = (\eta/\eta_0)$  and  $\theta = x$ , the equation assumes the form which can be dealt with by the present graphical method.

$$\text{To find } \sum_{n=1}^{\infty} (1/n^4).$$

This series occurs very frequently in physics and the result of the summation is often given without proof. The following proof is well known amongst mathematicians but it is given here so that it may be readily available to physics students. In addition, it supplies a useful example in the use of Fourier series. We have

$$\begin{aligned} \int_0^\pi x^4 \cos nx \, dx &= [x^4(\sin nx)/n]_0^\pi - (4/n) \int_0^\pi x^3 \sin nx \, dx \\ &= (-1)^n[(4\pi^3/n^2) - (24\pi/n^4)], \dots [\text{if } n \neq 0] \end{aligned}$$

after a few more reductions of a similar nature.

---

\* Ferguson, Phil. Mag. vol. xxxi. p. 37 (1916).

Similarly,  $\int_0^\pi x^2 \cos nx \, dx = (-1)^n (2\pi/n^2), \quad [\text{if } n \neq 0].$

$$\therefore \int_0^\pi (x^4 - 2\pi^2 x^2) \cos nx \, dx = (-1)^{n-1} (24\pi/n^4).$$

Now expand  $x^4 - 2\pi^2 x^2$  in a Fourier series for the range  $-\pi < x < \pi$ . We have, in the usual way,

$$x^4 - 2\pi^2 x^2 = a_0 + a_1 \cos x + a_2 \cos 2x + \dots + a_n \cos nx + \dots,$$

the "sine" terms being omitted since the function to be represented by the series is an even one.

$$\text{Now,} \quad a_0 = \frac{1}{2\pi} \int_{-\pi}^\pi (x^4 - 2\pi^2 x^2) \, dx = -\frac{7}{15} \pi^4.$$

$$\text{Also} \quad \int_{-\pi}^\pi x^4 \cos nx \, dx = 2 \int_0^\pi x^4 \cos nx \, dx = 2 \left[ (-1)^n \left( \frac{4\pi^3}{n^2} - \frac{24\pi}{n^4} \right) \right],$$

$$\text{and} \quad -2\pi^2 \int_{-\pi}^\pi x^2 \cos nx \, dx = (-1)^n \frac{8\pi^3}{n^2},$$

$$\therefore \int_{-\pi}^\pi (x^4 - 2\pi^2 x^2) \cos nx \, dx = (-1)^n \left( \frac{-48\pi}{n^4} \right),$$

$$\therefore a_n = \frac{1}{\pi} \int_{-\pi}^\pi (x^4 - 2\pi^2 x^2) \cos nx \, dx = (-1)^n \left( \frac{-48}{n^4} \right),$$

$$\therefore x^4 - 2\pi^2 x^2 = -\frac{7}{15} \pi^4 + 48 \sum_{n=1}^\infty (-1)^{n-1} \frac{\cos nx}{n^4}, \quad [-\pi < x < \pi].$$

Put  $x=0$ . Then

$$\sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^4} = \frac{7}{15 \cdot 48} \pi^4.$$

$$\begin{aligned} \text{But } \sum_{n=1}^\infty \frac{1}{n^4} &= \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^4} + 2 \left[ \frac{1}{2^4} + \frac{1}{4^4} + \frac{1}{6^4} + \dots + \frac{1}{(2n)^4} + \dots \right] \\ &= \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^4} + \frac{1}{2^3} \left[ 1 + \frac{1}{2^4} + \frac{1}{3^4} + \dots + \frac{1}{n^4} + \dots \right]. \end{aligned}$$

$$\therefore \frac{7}{8} \sum_{n=1}^\infty \frac{1}{n^4} = \sum_{n=1}^\infty (-1)^{n-1} \frac{1}{n^4} = \frac{7}{15 \cdot 48} \pi^4,$$

$$\therefore \sum_{n=1}^\infty \frac{1}{n^4} = \frac{\pi^4}{90}.$$

#### *On Sentis' Method of Determining the Surface Tension of a Liquid.*

As originally described by Sentis\* this method for the determination of the surface tension of a liquid requires a clean capillary tube drawn out to a fine jet. The tube is dipped into the liquid under investigation and then withdrawn: when it is held in a vertical position some of the liquid which has entered the tube exudes and forms a small spherical drop at the drawn-out end of the tube. The author has shown in a

---

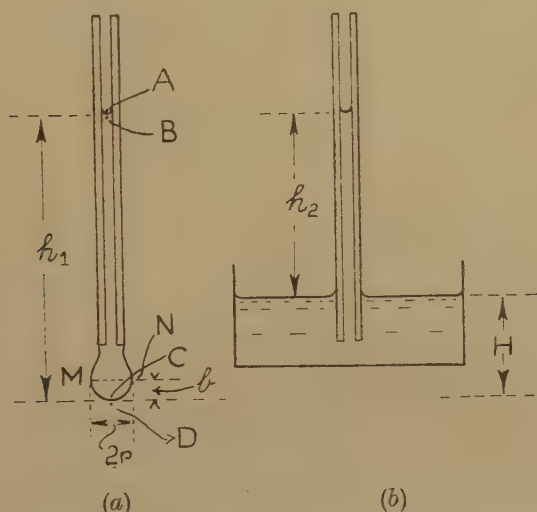
\* Sentis, *J. de Physique*, vol. vi. p. 571 (1887).



previous paper\* that considerable advantages accrue if the lower end of the tube is made considerably wider than its capillary portion. In fact, there is no inherent reason why a tube with an internal diameter of 1 mm. to 3 mm. throughout its whole length should not be used. In each case, however, it is no longer justifiable to assume that the lower portion of the pendent drop is hemispherical and the following analysis shows how a correction may be made. The argument is similar to one already used by Ferguson† in connection with the rise of a liquid in a capillary tube.

Fig. 4 (a) is a diagrammatic representation of the pendent drop and the liquid within a vertical capillary tube. Let  $P_0$  be the atmospheric pressure. Then at B, a point just below the liquid surface, A, in the

Fig. 4.



Theory of Sentsis' method of determining the surface tension of a liquid.

capillary tube, the pressure is  $P_0 - (2\gamma/R_0)$ , where  $R_0$  is the radius of curvature of the liquid surface at A, and  $\gamma$  is the surface tension of the liquid. Let MN be the trace of the plane of maximum cross-sectional area in the drop, while  $b$  is the height of this plane above D, the lowest point in the drop. Then the pressure at any point in this plane, and in the liquid, is  $P_0 - (2\gamma/R_0) + (h_1 - b)g\rho$ , where  $h_1 = AD$ ,  $g$  is the intensity of gravity and  $\rho$  is the density of the liquid. The force on MN, acting downwards and due to the above pressure, is  $\pi r^2[P_0 - (2\gamma/R_0) + (h_1 - b)g\rho]$ , where  $MN = 2r$ .

Consider now the equilibrium of the lower portion of the drop, which is assumed to be half an ellipsoid of revolution. The weight of this portion of the drop is  $(2/3)\pi r^2 b g \rho$ , and this force acts vertically downwards. The force due to surface tension is  $2\pi r \gamma$ , acting vertically upwards, while the

\* C. J. Smith, *Phil. Mag.* vol. xxxiii. p. 775 (1942).

† Ferguson, 'Manchester Memoirs,' vol. lxxv. No. 4, p. 11 (1921).

air exerts a force  $P_0 \cdot \pi r^2$ , which acts upwards. Hence for equilibrium,

$$2\pi r\gamma = \pi r^2 [-(2\gamma/R_0) + (h_1 - b)g\rho] + \frac{2}{3}\pi r^2 b g\rho.$$

Suppose now that a wide vessel containing the liquid is placed below the tube and the position of the wide vessel adjusted until the liquid in the capillary tube returns to its original position. Let  $h_2$  be the height of the liquid in the capillary above the free surface of the liquid in the wide vessel—cf. fig. 4 (b). Then  $(2\gamma/R_0) = g\rho h_2$ , and this is an exact equation. The condition for equilibrium may therefore be written

$$2\pi r\gamma = \pi r^2 [\{\overline{h_1 - h_2} - b\}g\rho] + \frac{2}{3}\pi r^2 b g\rho.$$

Calling  $h_1 - h_2 = H$  and writing  $a^2 = \gamma/(g\rho)$ , we get

$$2a^2 = (H - b)r + \frac{2}{3}rb = Hr - \frac{1}{3}br. \quad \dots \quad (1)$$

Now let  $R$  be the radius of curvature of the drop at its lowest point. Then consideration of the pressures at the points A, B, C and D shows that  $(2\gamma/R + 2\gamma/R_0) = g\rho h_1$ , and this is also an exact equation. Since  $(2\gamma/R_0) = g\rho h_2$  and  $h_1 - h_2 = H$ , we have

$$(2\gamma/R) = g\rho(h_1 - h_2) = g\rho H,$$

or

$$2a^2 = RH.$$

It is well known, however, that  $R = (r^2/b)$ , where  $2r$  is the maximum diameter of the pendent drop, *i. e.*,  $r$  is the semi-major axis of the ellipse which is the curve assumed to be the profile of the lower part of the drop. Hence  $b = (r^2 H / 2a^2)$ . Substituting this value for  $b$  in equation (1), we get

$$2a^2 = rH - (r^3 H / 6a^2), \quad \text{or} \quad 12a^4 - 6rHa^2 + r^3 H = 0.$$

Solving this quadratic in  $a^2$ , we get

$$a^2 = \frac{1}{4}[rH \pm rH\{1 - \frac{4}{3}(r/H)\}]^{\frac{1}{2}},$$

and only the positive sign is to be retained since, when  $r \rightarrow 0$ , we must have  $2a^2 = rH$ . Expanding the surd and rejecting terms in  $r^4$ , etc., we get

$$a^2 = \frac{1}{2}rH - \frac{1}{6}r^2 - \frac{1}{18} \frac{r^3}{H} = \frac{\gamma}{g\rho}.$$

$$\therefore \gamma = \frac{1}{2}g\rho \left[ rH - \frac{1}{3}r^2 - \frac{1}{9} \frac{r^3}{H} \right].$$

A convenient set-up for carrying out a determination of the surface tension of a liquid by Sentsis' method is shown in fig. 5. The wide glass vessel A, containing some of the liquid under investigation, is fixed to the head of a spherometer, the fixed legs of which rest on a piece of plate glass P, through which a hole has been bored to permit the passage of the central leg of the instrument. In this way the distance  $H$  may be determined directly, by observing the spherometer reading when the liquid in A just touches the pendent drop, and again when the meniscus in the capillary tube is restored to its former position by raising the vessel A: this act is accomplished by rotating the spherometer head. A vernier microscope is used to measure the maximum diameter of the pendent drop, and a second microscope is focussed on the liquid surface in the capillary tube. No further details are necessitated here.

The following result was obtained by this method for water :—

$$2r=0.3975 \text{ cm. Temp.}=21.0^{\circ} \text{ C.}$$

$$H=0.824 \text{ cm. } \rho=0.998 \text{ gm. cm.}^{-3}.$$

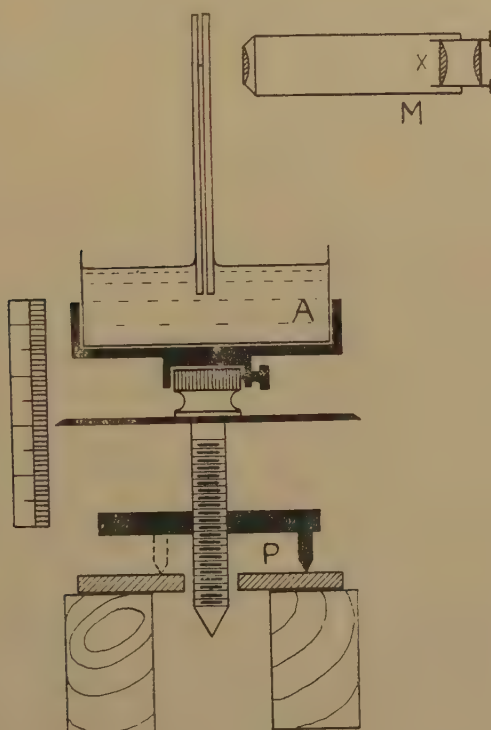
$$\begin{aligned} \gamma &= g\rho \left[ \frac{1}{2}rH - \frac{1}{6}r^2 - \frac{1}{18} \frac{r^3}{H} \right] \\ &= 981 \times 0.998 \left[ \frac{1}{2} \times 0.1988 \times 0.824 - \frac{1}{6} (0.1988)^2 - \frac{1}{18} \times \frac{(0.1988)^3}{0.824} \right] \\ &= 72.3 \text{ dyne. cm.}^{-1}. \end{aligned}$$

The surface tension of fleuss oil was also measured by this method with the following result :—

$$\text{Mean radius of drop}=0.119 \text{ cm. } H=0.6343 \text{ cm. Temp.}=25.0^{\circ} \text{ C.}$$

$$\text{Thus } \gamma=30.5 \text{ dyne. cm.}^{-1}.$$

Fig. 5.



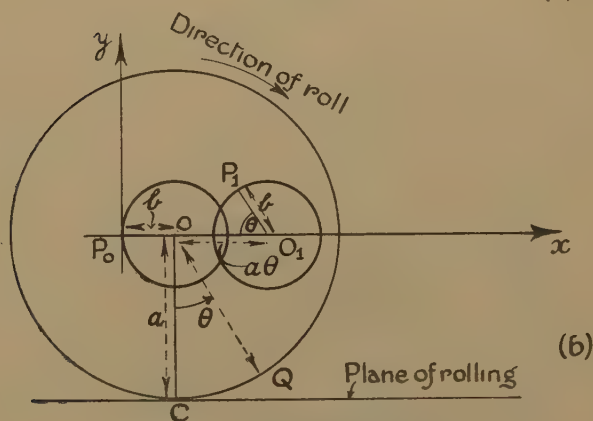
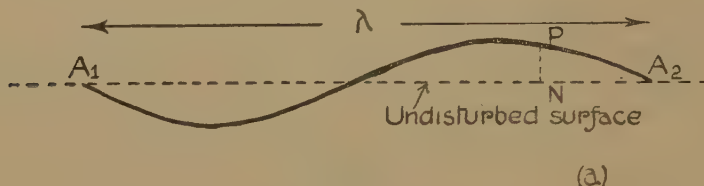
An attempt to measure the surface tension of glycerol by this method failed on account of the exceptionally high coefficient of viscosity which this liquid has at room temperature.

### *On a Simple Method of calculating the Velocity of Propagation of Ripples on a Liquid Surface*

The velocity  $c$  with which long plane waves travel across a liquid surface is known to be given by the equation  $c=\sqrt{(g\lambda/2\pi)}$ , where  $g$  is the

intensity of gravity, and  $\lambda$  is the wave-length of the disturbance. The profile of a portion of such a wave is shown in fig. 6 (a). Let P be a point on this profile and N its projection on the undisturbed surface of the liquid. If gravity were the only force acting, the increase in pressure at N due to the wave would be  $g\rho \cdot PN$ , where  $\rho$  is the density of the liquid. Surface tension will give rise to an additional pressure  $\gamma r^{-1}$ , where  $\gamma$  is the surface tension of the liquid and  $r$  is the radius of curvature at P. If the amplitude is small compared with  $\lambda$ , the profile of the wave may be regarded as the curve traced out by a point near to the centre of a circle rolling along a plane surface. Let  $a$  be the radius of this circle and  $b$  the distance of P from its centre O—cf. fig. 6 (b). Let the position

Fig. 6.



Velocity of ripples on a liquid surface.

of P be given with reference to a system of fixed rectangular axes  $P_0x$  and  $P_0y$ , where  $P_0O$  is parallel to the plane of rolling. Let time be measured from the instant when P is at  $P_0$ . Let  $\theta$  be the angle of roll: then  $\angle QOC = \theta$ , where the point Q will be the point of contact which the rolling circle makes with the plane of rolling when  $P_0$  has moved to  $P_1$ , where  $\angle P_1O_1P_0 = \theta$  and  $OO_1 = a\theta$ . The instantaneous coordinates of  $P_1$  are then

$$x = b + a\theta - b \cos \theta = a[\theta + (b/a) - (b/a) \cos \theta]$$

$$= a\theta, \quad [\text{if } (b/a) \rightarrow 0]$$

and

$$y = b \sin \theta.$$

$$\therefore \frac{dy}{dx} = \frac{b}{a} \cos \theta, \quad \text{and} \quad \frac{d^2y}{dx^2} = -\frac{b \sin \theta}{a^2}.$$



If  $\frac{dy}{dx}$  is small, a condition appropriate to the present problem, the radius of curvature is  $\left(\frac{d^2y}{dx^2}\right)^{-1}$ . Thus

$$r = \left(\frac{d^2y}{dx^2}\right)^{-1} = -\left(\frac{b \sin \theta}{a^2}\right)^{-1} = -\frac{a^2}{PN}.$$

Hence, neglecting the negative sign, in the case of a wave the excess pressure due to surface tension is

$$\frac{\gamma}{r} = \gamma \cdot \frac{PN}{a^2}.$$

But  $2\pi a = \lambda$ , so that the above excess pressure may be put equal to  $(4\pi^2 \cdot PN \cdot \gamma / \lambda^2)$ . Now when the pressure at N is  $g\rho \cdot PN$ , the velocity is  $\sqrt{(\rho\lambda/2\pi)}$ . Therefore, when the pressure at N is  $[g\rho + (4\pi^2\gamma/\lambda^2)]PN$ , the velocity will be given by

$$c = \sqrt{\frac{g\lambda}{2\pi} \left(1 + \frac{4\pi^2\gamma}{\lambda^2 g\rho}\right)} = \sqrt{\frac{g\lambda}{2\pi} + \frac{2\pi\gamma}{\lambda\rho}}.$$

It seems justifiable to do this since we may regard the effect as due to an increase in  $g\rho$  in the ratio  $[1 + (4\pi^2\gamma/\lambda^2 g\rho)] : 1$ .

#### *Note on the Calibration of Capillary Tubes used in Viscometers.*

Let us assume that a capillary tube of length  $l$  is divided into  $n$  portions, the length of the  $k$ th section being  $(\Delta l)_k$ —cf. fig. 7 (a). Let  $r_k$  be the mean radius of this section. Let  $p_1$  and  $p_2$  be the pressures at the entrance and exit ends of the capillary tube, while  $p'$ ,  $p''$ , ... are the pressures at the exit ends of the first, second, ... sections. Then if  $v$  is the volume of liquid flowing steadily in  $t$  seconds through the tube, we have, if  $r$  is the effective radius of the tube,

$$\begin{aligned} v &= \frac{\pi r^4 (p_1 - p_2) t}{8\eta l} \\ &= \frac{\pi r_1^4 (p_1 - p') t}{8\eta (\Delta l)_1} = \frac{\pi r_2^4 (p' - p'') t}{8\eta (\Delta l)_2} = \dots = \frac{\pi r_k^4 (p^{k-1} - p^k) t}{8\eta (\Delta l)_k} = \dots \\ \therefore v &= \frac{\pi t}{8\eta} \frac{(p_1 - p')}{\frac{(\Delta l)_1}{r_1^4}} = \frac{\pi t}{8\eta} \frac{(p' - p'')}{\frac{(\Delta l)_2}{r_2^4}} = \dots = \frac{\pi t}{8\eta} \frac{(p^{k-1} - p^k)}{\frac{(\Delta l)_k}{r_k^4}}, \\ \therefore v &= \frac{\pi t}{8\eta} \frac{p_1 - p' + p' - p'' + \dots + p^{k-1} - p^k + \dots}{\sum_{k=1}^n \frac{(\Delta l)_k}{r_k^4}} \\ &= \frac{\pi t}{8\eta} \frac{(p_1 - p_2)}{\sum_{k=1}^n \frac{(\Delta l)_k}{r_k^4}}. \end{aligned}$$

This explains why it is necessary to consider the expression  $\sum_{k=1}^n \frac{(\Delta l)_k}{r_k^4}$ , when the tube is not uniform in section. A similar argument may be

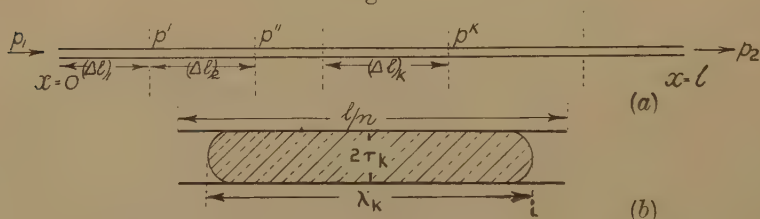
applied to the case of a gas flowing steadily through a capillary tube. It gives for  $m$ , the mass of gas flowing in  $t$  seconds, at absolute temperature  $T$ , if  $\mathcal{R}$  is the gas constant per gramme of gas,

$$m\mathcal{R}T = \frac{\pi}{16\eta l} \frac{(p_1^2 - p_2^2)}{\sum_{k=1}^n \frac{(\Delta l)_k}{r_k^4}},$$

so that the same summation is required.

To carry out the calibration the capillary tube is divided into  $n$  equal lengths: the actual length of each section being 4 or 5 cm. Let  $M$  be the mass of mercury required to fill the tube completely—this is estimated from a knowledge of the mass and length of a mercury pellet which almost fills the tube. Then let a mercury pellet of length less than  $(l/n)$  be introduced into the tube and let its length be measured in each section of the tube. For the  $k$ th section—*cf.* fig. 7 (b)—let  $\lambda_k$  be the distance indicated, *i. e.* the length of the pellet. Then  $\pi\rho r_k^2\lambda_k$ , where  $\rho$  is

Fig. 7.



Calibration of a capillary tube for use in a viscometer.

the density of mercury at the temperature of the tube, is very nearly equal to the mass of the mercury pellet. We may therefore write

$$\pi\rho r_k^2\lambda_k = C_k,$$

and the question arises as to whether or not  $C_k$  varies according to the section of the tube which the pellet occupies. Strictly speaking it does vary, because the radius of the curved portion at either end varies with the position of the pellet in the tube. This variation, however, is small if the tube is approximately uniform in diameter, and so it is only for such tubes that the method of calibration is applicable. In such an instance we may write  $C_k = C$ , a constant, and then we have

$$\begin{aligned} M &= \pi\rho \sum_{k=1}^n \left( r_k^2 \frac{l}{n} \right) = \pi\rho \frac{l}{n} \sum_{k=1}^n (r_k^2) \\ &= \frac{l}{n} C \sum_{k=1}^n \left( \frac{1}{\lambda_k} \right). \end{aligned}$$

$$\therefore \frac{1}{C} = \frac{l}{nM} \sum_{k=1}^n \left( \frac{1}{\lambda_k} \right).$$

But

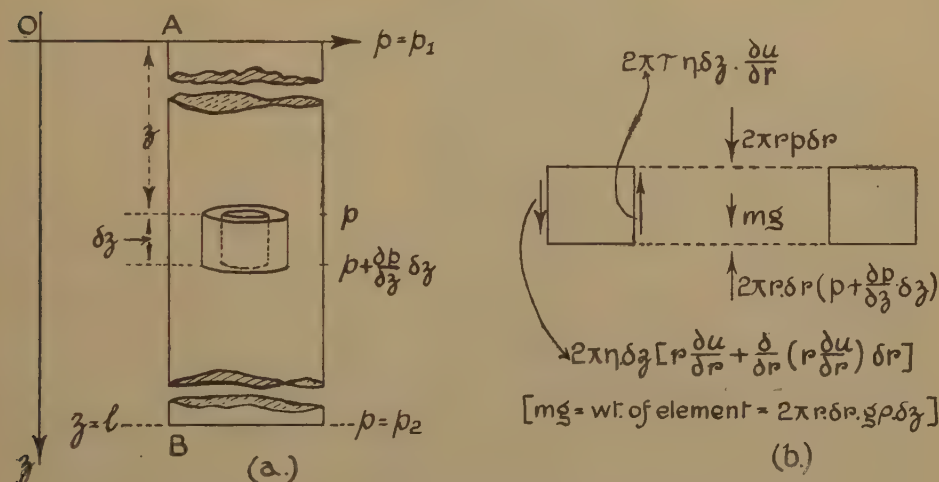
$$\begin{aligned} \frac{1}{r_k^4} &= \frac{\pi^2 \rho^2 \lambda_k^2}{C^2}, \\ \therefore \sum_{k=1}^n \frac{(\Delta l)_k}{r_k^4} &= \frac{l}{n} \sum_{k=1}^n \left( \frac{1}{r_k^4} \right) \quad \left[ \because (\Delta l)_k = \frac{l}{n} \right] \\ &= \frac{l}{n} \frac{\pi^2 \rho^2 \Sigma (\lambda_k^2)}{C^2} \\ &= \frac{l^3}{n^3} \frac{\pi^2 \rho^2}{M^2} \sum_{k=1}^n (\lambda_k^2) \cdot \left[ \sum_{k=1}^n \left( \frac{1}{\lambda_k} \right) \right]^2. \end{aligned}$$

This method of calibrating a capillary tube has been employed in the experiment which follows.

*On the Flow of an Incompressible Liquid through a Vertical Capillary Tube, and a Constant-head Viscometer in which use is made of a Vertical Capillary Tube.*

Let AB, fig. 8 (a), be the vertical tube through which the liquid is flowing steadily. Consider the forces on the small cylindrical element

Fig. 8.



Steady flow of a liquid through a vertical capillary tube.

whose ends are at distances  $z$  and  $(z + \delta z)$  from A, while the radii of its curved surfaces are  $r$  and  $(r + \delta r)$ . Let  $u$  be the velocity in the liquid at a point on the inner curved surface of the above element. At a point on the outer curved surface the velocity will be  $\left( u + \frac{\partial u}{\partial r} \delta r \right)$ . The forces acting on the element are shown in fig. 8 (b), where it must be remembered that the forces due to the viscosity of the liquid are actually distributed over the curved surfaces. Likewise the forces arising from the pressure in the liquid are distributed over the flat ends of the element.

In the steady state the resultant force downwards is zero since the liquid is moving without acceleration. Hence

$$-2\pi r\eta \delta z \cdot \frac{\partial u}{\partial r} + 2\pi\eta \delta z \left[ r \frac{\partial u}{\partial r} + \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right) \delta r \right] \\ + 2\pi r \delta r \cdot g\rho \delta z + 2\pi r p \delta r - 2\pi r \cdot \delta r \left( p + \frac{\partial p}{\partial z} \delta z \right) = 0,$$

where  $g$  and  $\rho$  have their usual meanings.

$$\therefore g\rho r \cdot \delta r + \eta \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right) \delta r - r \delta r \cdot \frac{\partial p}{\partial z} = 0.$$

$$\text{Now} \quad \frac{\partial p}{\partial z} = - \left( \frac{p_1 - p_2}{l} \right),$$

$$\text{so that} \quad \left( g\rho + \frac{p_1 - p_2}{l} \right) r = -\eta \frac{\partial}{\partial r} \left( r \frac{\partial u}{\partial r} \right).$$

$$\therefore \left( g\rho + \frac{p_1 - p_2}{l} \right) \frac{r}{2} + \frac{A}{r} = -\eta \frac{\partial u}{\partial r}, \quad [A = \text{integ. const.}]$$

$$\therefore \left( g\rho + \frac{p_1 - p_2}{l} \right) \frac{r^2}{4} + A \log r + B = -\eta u. \quad [B = \text{integ. const.}]$$

But  $A=0$ , since  $u \neq \infty$  when  $r \rightarrow 0$ , and  $r=a$  gives  $u=0$ ,

$$\therefore B = -[g\rho + (p_1 - p_2)/l](a^2/4), \\ u = [g\rho + (p_1 - p_2)/l][(a^2 - r^2)/4\eta].$$

If  $V$  is the volume of liquid flowing per second through the tube, we have

$$V = \int_0^a u \cdot 2\pi r \cdot dr = [\pi a^4 (p_1 - p_2 + g\rho l)/8\eta l],$$

after the integration has been performed.

#### *A Vertical Viscometer for Liquids.*

A viscometer to which the above theory is immediately applicable is shown in fig. 9. It is designed for students' use primarily and the minor changes necessary to make it a precision instrument will be apparent. A wide glass tube  $G$  contains the liquid under investigation and the capillary tube  $AB$  is carried in a vertical position by a rubber bung fitted to the lower part of  $G$ . A second rubber bung fitted into the upper end of the tube  $G$  carries the glass tubes  $C$ ,  $D$  and  $E$ , together with the thermometer  $T_1$ . Liquid is introduced, when necessary, through  $E$ , the rubber cap on  $D$  being removed so that the displaced air may escape from the apparatus. The tube  $E$  also permits air to be blown through the liquid in order that it shall be thoroughly stirred before observations are made. With  $D$  closed, as liquid exudes from the capillary tube air enters the apparatus through  $C$  so that the pressure at the lower end of  $C$  is almost constant and the effect of this slight variation may be eliminated by the method adopted below.

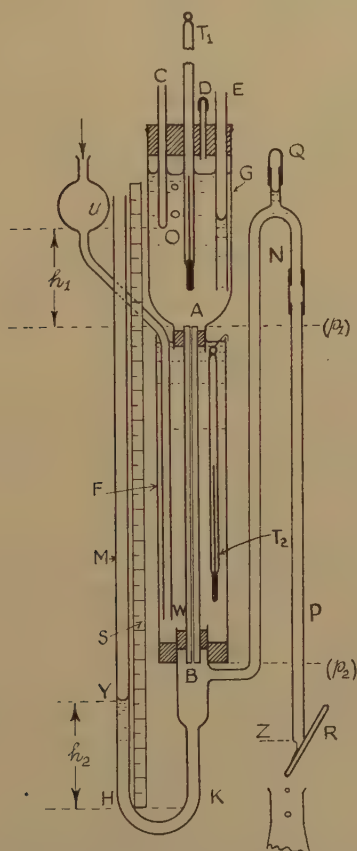
Now if drops of liquid are allowed to fall from  $B$  directly into the air the pressure at  $B$  within the liquid is not constant. To overcome this difficulty the lower end of the capillary tube  $AB$  is fitted into a glass



tube of the shape indicated : this is connected to a tube M which serves as a manometer while the inverted U-tube BNZ to which it is connected provides a means of escape for the liquid. The tube Q attached to the top of the U-tube permits air bubbles to escape during the process of filling the viscometer : when this is in use Q is closed with a rubber cap.

In order to ensure steady conditions at the end of the tube Z, a short glass rod R is held rigidly against Z and the drops falling from R are collected in a weighed flask.

Fig. 9.



A vertical viscometer for use with liquids at room-temperature.

Let  $P$  be the atmospheric pressure. Then in the liquid at the lower end O of the tube C the pressure is  $(P - \alpha)$ , where  $\alpha$  represents the reduction in pressure due to surface tension and a correction due to the fact that the surface of the bubble is not fixed in position. Thus  $\alpha$  is variable, strictly, but probably assumes a constant mean value over the time required for a conveniently measurable quantity of liquid to escape. Hence

$$p_1 = P - \alpha + g\rho h_1,$$

where  $h_1$  is the depth of A below O, and  $g$  and  $\rho$  have their usual meanings. Now at B the pressure is  $p_2$ , where

$$p_2 = P + g\rho h_2 + \beta,$$

where  $\beta$  is a correcting constant due to the fact that the zero from which  $h_2$  is measured is not at B—it is convenient to have a scale in cm., etc., S, resting on the lower portion of the tube HK—and also necessary in part because the pressure in the liquid at Y is not quite atmospheric.

The formula established in the previous section, therefore, may be written

$$\begin{aligned} V &= [(\pi a^4)/(8\eta l)][P - \alpha + g\rho h_1 - P - g\rho h_2 - \beta + g\rho l] \\ &= [(\pi a^4)/(8\eta l)][C - g\rho h_2], \end{aligned}$$

where C is a constant for a given apparatus. If we make a series of observations using different lengths of the tube NZ, we may construct the straight-line graph  $x=h_2$ ,  $y=V$ . The slope of this line is  $-[(\pi a^4)/(8\eta l)]g\rho$ , so that  $\eta$  may be found. If the tube AB is not quite uniform, the slope will be

$$- \left[ (\pi g \rho) / \left\{ 8\eta \Sigma \left( \frac{\Delta l}{r^4} \right) \right\} \right].$$

Since the apparatus cannot conveniently be set up in a thermostat it is essential to work in a room where the temperature is fairly constant. It is also necessary to surround the capillary tube by a wide glass tube F containing water. This is stirred by forcing air through the glass tube UW, the bulb at U preventing water from entering the apparatus in which air under pressure is produced. A mercury thermometer  $T_2$  gives the temperature of the capillary tube.

In collecting the drops it is advisable to place the flask in position immediately after a drop has fallen from the end of R and to remove it after a time interval when another drop has just fallen. The time between the fallings of these two drops must be measured with an accurate stop-watch.

The following result was obtained for distilled (air-free) water at  $23.0^\circ\text{C}$ . It will be seen that  $h_2$  was varied from 18.55 cm. to 52.70 cm., *i.e.* the position of Y varied from one near H to one near A. The temperature changed from  $22.9^\circ\text{C}$ . to  $23.1^\circ\text{C}$ . during the experiment.

Mass of water [gm.]	Time of flow [sec.]	Rate of flow [gm. sec. <sup>-1</sup> ]	$h_2$ [cm.]
13.723	604.1	0.0227	18.55
13.651	604.6	0.0226	18.55
3.600	610.5	0.0059	52.70
3.475	603.6	0.0058	52.70
12.240	605.0	0.0202	23.80
10.682	607.0	0.0176	29.20
8.568	602.1	0.0142	34.95
6.730	603.0	0.0112	41.00

These give  $\eta = (9.43 \pm 0.07) \times 10^{-3}$  gm. cm.<sup>-1</sup> sec.<sup>-1</sup> at  $23.0^\circ\text{C}$ .

*Note on the Theory of Ostwald Viscometers.*

It is surprising to find that in describing the use of his well-known viscometer for the comparison of the viscosities of liquids at constant temperature Ostwald, together with many others, assumes Poiseuille's formula for the flow of a liquid through a horizontal tube to be applicable when the tube is vertical. As long as the instrument is used to compare viscosities the final result is unaffected, but it seems worth while to put on record the correct theory. Fig. 10 (a) is a diagram of an Ostwald viscometer in which, for the sake of generality, it is assumed that the upper and lower bulbs are not identical: the cross-section A of the upper bulb A is a function of  $x$  only, *i. e.*  $A=\phi(x)$ , while for B we may similarly write  $B=\psi(y)$ , where  $x$  is the vertical distance through which the meniscus in A falls in time  $t$ , and  $y$  is the distance through which the meniscus in B moves upward in the same time. If  $V$  is the volume of liquid which passes through a vertical capillary tube in time  $t$ , it has been shown above that

$$V=[\pi a^4(p_1-p_2+g\rho l)/8\eta l]t,$$

where the symbols have their earlier significance, so that

$$dV=[\pi a^4(p_1-p_2+g\rho l)/8\eta l]dt.$$

But if  $dx$  is the fall in A in time  $dt$ , and  $dy$  the corresponding rise in B,  $A dx=dV=B dy$ . If  $h_0$  and  $H_0$  are the vertical distances as indicated in the diagram,

$$\begin{aligned} A dx &= [(\pi a^4 g \rho)/(8\eta l)] [h_0 - x - H_0 - y + l] dt \\ &= k(\rho/\eta) [K - x - y] dt \quad [k \text{ and } K \text{ are constants}] \\ &= k(\rho/\eta) f(x) dt, \end{aligned}$$

where  $f(x)$  may be written for  $(K - x - y)$ , which is a function of  $x$  only since

$$dV = \phi(x) dx = \psi(y) dy,$$

and

$$\therefore V = \int_{t=0}^t dV = \int_0^x \phi(x) dx = \int_0^y \psi(y) dy.$$

Thus  $y$  is a function of  $x$  only.

Hence 
$$\frac{A dx}{f(x)} = k(\rho/\eta) dt.$$

$\therefore$  If  $\tau$  is the time required for the level of the liquid in A to fall from its initial position to a level  $x=X$ , say (shown by a mark on the stem of the instrument),

$$\int_0^X \frac{A dx}{f(x)} = k(\rho/\eta) \tau.$$

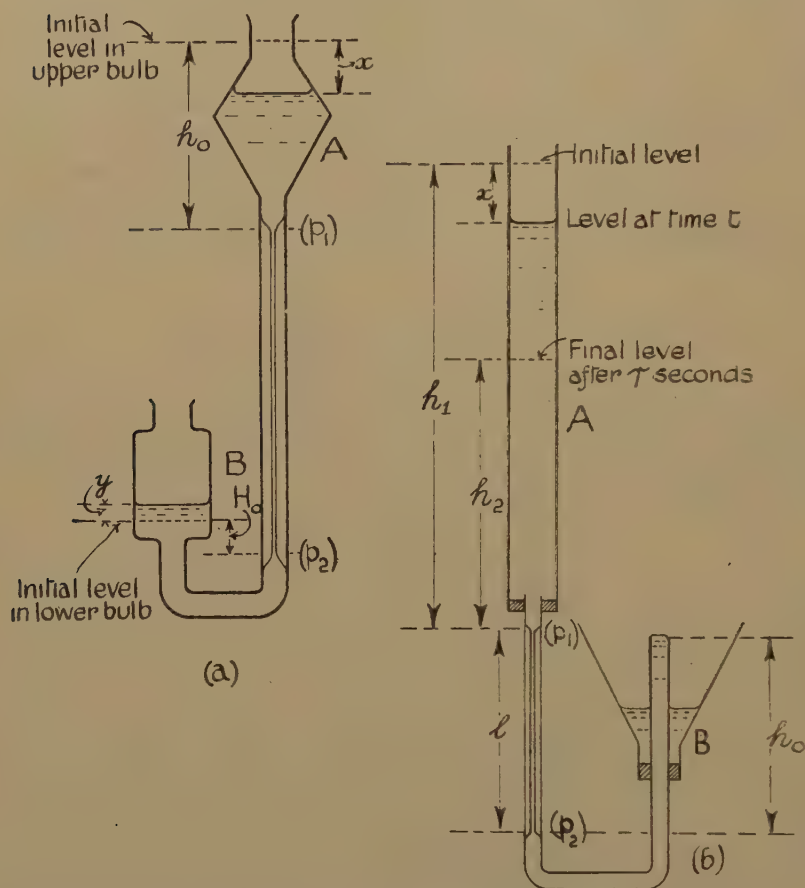
Now for a given viscometer at constant temperature  $\int_0^X \frac{A dx}{f(x)}$  is constant

and therefore under such conditions  $\rho\tau\eta^{-1}$  is invariable, so that if two liquids are used in turn in the viscometer

$$(\eta_1/\eta_2) = (\rho_1\tau_1)/(\rho_2\tau_2).$$

Another viscometer, for which the usual theory is not correct, is the type used by Ostwald and Auerbach \* in their work on colloidal solutions. Mukoyama † has given a formula for this viscometer: Hatschek ‡ has shown this to be wrong but even this author has failed to obtain a correct formula, and a still more recent author Philippoff §, who refers to the

Fig. 10.



Theory of two types of Ostwald viscometer.

work of Hatschek, has apparently not noticed the error. The formula now suggested may be derived as follows. Fig. 10 (b) shows in section an Ostwald and Auerbach viscometer. It will be assumed that the wide

\* Ostwald and Auerbach, *Koll. Zeitsch.* vol. xli. p. 56 (1927).

† Mukoyama, *Koll. Zeitsch.* vol. xli. p. 62 (1927).

‡ Hatschek, 'The Viscosity of Liquids,' p. 26 (Bell & Sons, 1928).

§ Philippoff, 'Viskosität der Kolloide,' p. 66 (Dresden, 1942).



tube A is uniform in cross-section : let  $x$  be the distance through which the liquid meniscus in it falls in time  $t$ . Then  $p_1$  and  $p_2$ , the pressures at the upper and lower ends of the capillary tube, are given by  $p_1 = P + g\rho(h_1 - x)$ , and  $p_2 = P + g\rho h_0$ , where  $h_1$  and  $h_0$  are the distances indicated,  $P$  is the atmospheric pressure, and all corrections for surface tension effects are neglected. Hence

$$dV = [(\pi a^4)/(8\eta l)]g\rho[h_1 - x - h_0 + l],$$

and since  $dV = A_0 dx$ , if  $A_0$  is the constant cross-sectional area of the tube A, we have

$$A_0 dx = [(\pi a^4)/(8\eta l)]g\rho[\kappa - x]dt,$$

where  $\kappa$  is a constant.

$$\therefore A_0 \int_0^{h_1 - h_2} \frac{dx}{\kappa - x} = \frac{\pi a^4}{8\eta l} g\rho \int_0^\tau dt,$$

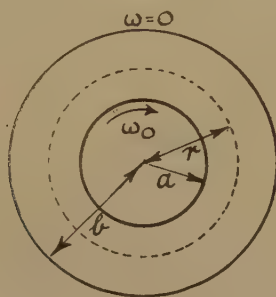
where  $h_2$  is the height of the liquid surface in A above the top of the capillary tube at time  $\tau$ .

$$\therefore A_0 \log \left[ (\kappa - x) \right]_{h_1 - h_2}^0 = \frac{\pi a^4}{8\eta l} g\rho \tau,$$

$$\text{i. e.} \quad A_0 \log \frac{h_1 - h_0 + l}{h_2 - h_0 + l} = \frac{\pi a^4}{8\eta l} g\rho \tau,$$

so that  $\eta$  may be obtained.

Fig. 11.



Theory of a rotating viscometer.

### *On the Couple exerted on a Cylinder placed in a Rotating Fluid.*

The theory of this important problem, as given in several textbooks, can only be described as slovenly. The reason for this remark is that the writers endeavour to interpret the calculus in a manner which was never intended by the mathematicians : but let us proceed.

The space between two coaxial cylinders, shown in section in fig. 11, is imagined to be filled with a viscous fluid. Let us assume that the outer cylinder is at rest, and that the inner cylinder can be caused to rotate with constant angular velocity  $\omega_0$  by means of a couple which

can be measured. Let the direction of the couple be clockwise and let  $a$  and  $b$  be the radii of the inner and outer cylindrical surfaces in contact with the fluid.

Consider the equilibrium of that part of the system lying within a cylinder of radius  $r$  and length  $l$ . The rate of shearing in the fluid at a point on the surface of this cylinder is  $r \frac{\partial \omega}{\partial r}$ , so that the tangential stress is  $\eta r \frac{\partial \omega}{\partial r}$ , where  $\eta$  is the coefficient of viscosity of the fluid. Hence the fluid outside the above cylinder exerts on it a force whose moment about the axis of rotation is  $2\pi r^3 \eta l \frac{\partial \omega}{\partial r}$ .

This is the magnitude of the couple acting in a clockwise direction which we have taken to be positive. Let  $\Gamma$  be the external applied couple. Then since there is no acceleration the total couple must be zero, *i. e.*

$$2\pi r^3 \eta l \frac{\partial \omega}{\partial r} + \Gamma = 0, \quad \text{or} \quad \frac{2\pi \eta l}{\Gamma} \omega = \frac{1}{2r^2} + A,$$

where  $A$  is an integration constant. At  $r=b$ ,  $\omega=0$ , and at  $r=a$ ,  $\omega=\omega_0$ , so that finally we obtain

$$\eta = \frac{\Gamma}{4\pi l \omega_0} \left[ \frac{1}{a^2} - \frac{1}{b^2} \right].$$

Now in the textbook method of obtaining this equation an attempt is made to use the fact that  $\frac{\partial \omega}{\partial r}$  is negative. This, as the method given above shows, is unnecessary and, furthermore, it is undesirable, as well as wrong, to insert what are equivalent to numerical values in any proof which is general in its scope.

*The Construction of a Steady Flow Calorimeter suitable for use in Laboratory determinations of J and the Specific Heat of Liquids.*

The purpose of this note is to describe the construction of a steady flow calorimeter for ordinary laboratory purposes. The flow-tube AB is made of glass—preferably pyrex—and has the dimensions shown in fig. 12. It is surrounded by a wide glass tube G, the ends of which are closed by rubber bungs. For convenience, one of these bungs, say D, is placed in position on the end B of the flow-tube before the end A is joined to the narrow central portion. It is advantageous, unless one is a skilled glass-blower, to make the final join at, say, K. The second bung C is cut into two portions, after it has been bored to fit A, and inserted after the wider tube has been placed in position.

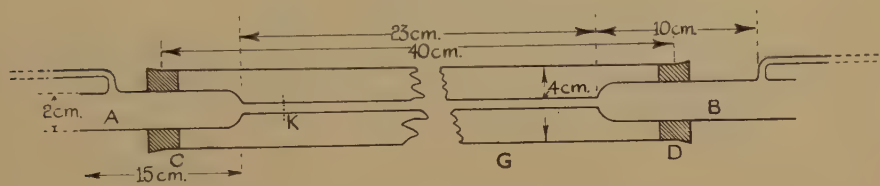
The wire in which energy is dissipated must be kept straight and not allowed to touch the glass, in order to avoid local "hot-spots" in the glass\*: if these exist the usual method of making a correction for the rate of heat exchange between the calorimeter and its surroundings

---

\* Barnes, Phil. Trans. A, vol. cxcix, p. 149 (1902).

becomes invalid. The wire is a piece of manganin, No. 28 S.W.G. and about 22 cm. long, so that its electrical resistance is approximately one ohm. The ends of this wire are first covered with a thin layer of "Easyflo" solder\*: this is a silver solder and must be used since manganin cannot be soldered successfully with ordinary lead-tin solder. As manganin rapidly oxidizes when red hot the following method must be used to coat the ends of the wire with the "Easyflo" solder. An iron pot, of the dimensions shown in fig. 13 (a), is made from an iron rod and filled with the solder, melted by means of a "Fisher" burner. The melt is covered with a layer of the special flux sold for use with this solder. The end of the wire to be coated is dipped into the molten solder, allowed to remain in it for a few seconds, and then withdrawn. The wire thus prepared is soldered into short pieces of thick copper wire, S.W.G. No. 18. These should have been screwed and one of them drilled to take a piece of copper wire, S.W.G. No. 24. Soft solder is used to fix this copper wire in position—cf. fig. 13 (b).

Fig. 12.



Glass components of a steady-flow calorimeter for laboratory determinations of J.

A brass cup of the form shown in fig. 13 (d) is also made. Its diameter is such that it just fits into A (fig. 12), and its outer surface has a deep spiral groove in it. It is screwed at one end so that the prepared manganin wire may be fitted to it. The screwed joint is soft-soldered. A second brass cup, similar to the above, must also be made: it differs from the first one in that its rounded end-portion is separated from the main portion, but it must be possible to screw them together to form a complete "cup." Sufficient details are given in the diagram to render further description unnecessary.

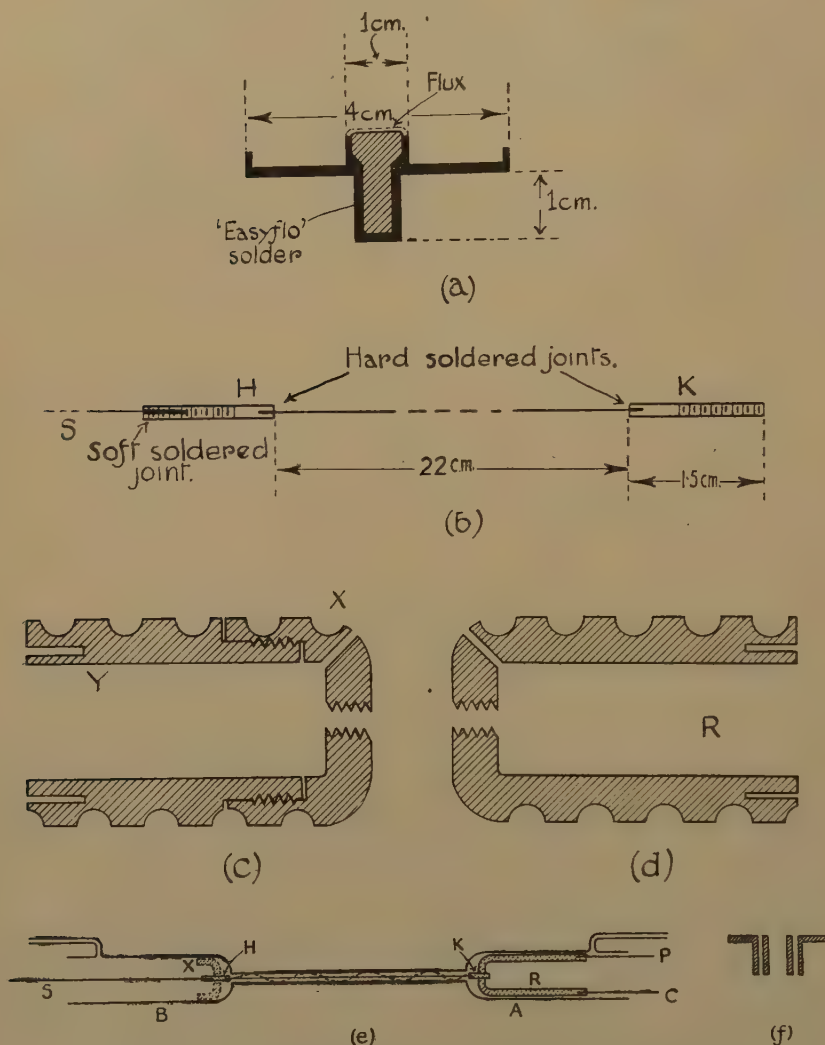
When the end K of the heater has been soldered to the cup R a piece of thick cord is wrapped round the manganin wire, held in position with loops of silk threads at its ends: this causes the liquid to be churned as it passes over the wire and also prevents local "hot-spots" from being formed.

The portion X of the second cup, with a blob of soft solder and some flux in a suitable position, is then passed over the copper wire S and screwed on to H. A slot in the flat end of X facilitates the carrying out

\* Obtainable from Messrs. Johnson Matthey & Co. Ltd., 78 Hatton Garden, London, E.C.1.

of this operation. The glass tube is then gently heated near to where X lies in position: the soft solder melts so that S may be withdrawn and when the tube has cooled X is held rigidly in position and the portion Y of the second cup may be added. Of course the potential leads P

Fig. 13.



Details of the construction of a steady-flow calorimeter.

and the current leads C must be in position before the above components are assembled.

Two ebonite "bungs"—a section of one is shown in fig. 13 (f)—are then made to fit the ends of the flow-tube: each is provided with a wide hole to take the mercury-in-glass thermometer and two narrower holes



through which pass the current and potential leads. Each ebonite bung is slipped into position, gently warmed, covered with an adequate layer of Chatterton's compound, and thus effectively sealed into position.

The whole can now be mounted in the wider tube and is then ready for filling with the calorimetric liquid. The calorimeter should be mounted in a vertical position for this purpose and the liquid allowed slowly to enter through the lower side tube. Small "air-vents" drilled in the ends of the cups permit air to escape from them during the process of filling. If air bubbles are troublesome and prevent the instrument from being filled, say, with water, it should be filled with methylated spirit (lower surface tension) and this liquid then removed by allowing water to flow rapidly through the calorimeter.

The rest of the "story" is so well known that it need not be repeated here. It is, of course, essential that the flow of liquid should be steady—at least to within one part in 500. As a preliminary a rough value for  $J$  should be calculated, no correction for heat-exchange being made. If this is about  $4.26 \text{ joule cal.}^{-1}$  the apparatus is probably in working order. It is found that good results are obtained if energy is dissipated at a rate of about 25 watt for the more rapid flow, which is necessary if the correction for heat-exchange is to be made: it is essential to do this if an accurate result is required.

*On the Specific Heat of a Gas at Constant Volume.—The Correction for the Expansion of the Vessel.*

Since the volume of the vessel increases during an experiment the gas, strictly speaking, is not maintained at constant volume. Consequently work is done by the gas and the thermal equivalent of this must be subtracted from the heat supplied to the sphere and its contents when writing down the equation from which the specific heat is deduced. Let  $V_1$ =initial volume,  $p_1$ =initial pressure, and  $T_1$ =initial temperature of the gas on the absolute scale. At any instant during the course of the experiment let  $p$  be the pressure,  $V$  the volume, and  $T$  the temperature. Then

$$(pV)/T = (p_1V_1)/T_1, \text{ or } p = [(p_1V_1)/V] \times [T/T_1],$$

and 
$$V = V_1(1 + \gamma t),$$

where  $\gamma$  is the coefficient of cubical expansion of the material of the sphere, and  $t$  denotes  $(T - T_1)$ . The well-known and more exact expression for  $V$  is not required.

$\therefore$  Work,  $W$ , done by the gas during the expansion of the containing vessel is given by

$$\begin{aligned} W &= \int_{V_1}^{V_2} p dV \quad [\text{the suffix 2 denotes conditions at the end of an experiment}] \\ &= p_1 \int_{T_1}^{T_2} \frac{T}{T_1} \{1 + \gamma(T - T_1)\}^{-1} \gamma V_1 dT = p_1 V_1 \int_0^{T_2 - T_1} (1 + \alpha t)(1 + \gamma t)^{-1} \gamma dt, \end{aligned}$$

where  $\alpha = T_1^{-1}$ , Joly \* takes  $\alpha$  to be the coefficient of increase in pressure at constant volume, *i. e.*  $T_0^{-1}$ , but this is not correct since the initial temperature is not  $0^\circ \text{C}$ . Hence

$$\begin{aligned} W &= p_1 V_1 \left[ \alpha(T_2 - T_1) + \frac{\gamma - \alpha}{\gamma} \log \{1 + \gamma(T_2 - T_1)\} \right] \\ &= p_1 V_1 \left[ \alpha(\theta_2 - \theta_1) + \frac{\gamma - \alpha}{\gamma} \log \{1 + \gamma(\theta_2 - \theta_1)\} \right], \end{aligned}$$

if temperatures measured on a centigrade scale are, in general, denoted by  $\theta$ . Joly obtains an expression for  $W$  which differs from this: it is in error probably because the limits of integration were confused.

The expression just obtained for  $W$ , the work done by the gas in the expansion, is exact, in so far as the gas is an ideal one. A good approximation is obtained by expanding the logarithm and neglecting powers of  $(\theta_2 - \theta_1)$  above the second. Then

$$W = p_1 V_1 [\gamma(\theta_2 - \theta_1) \{1 + \frac{1}{2}(\alpha - \gamma)(\theta_2 - \theta_1)\}].$$

This does not differ from Joly's final expression for  $W$ , except that he interprets  $\alpha$  incorrectly.

#### *On the Difference between the two Principal Specific Heats of an Ideal Gas.*

The usual elementary proof of this well-known equation leaves much to be desired, in that an ideal gas is not defined correctly and a clear statement about what happens to the energy supplied when heating occurs is not made. Let us therefore consider one gram of an ideal gas. Now such a gas has for its characteristic equation  $pv = \mathcal{R}T$  and, in addition, its internal energy is a function of the temperature only and is entirely kinetic, *i. e.* the Joule-Thomson cooling-effect is zero.

Let  $c_v$  be the specific heat of the gas at constant volume, so that when one gram of the gas is heated at constant volume the heat required to raise the temperature  $1^\circ$ , say from  $T_1$  to  $T_2$  on the absolute scale, is  $c_v$  cal. This energy is spent entirely in increasing the kinetic energy of the molecules. Now let the one gram of gas be restored to its original state and then be heated at constant pressure, the volume changing from  $v_1$  to  $v_2$  and the temperature from  $T_1$  to  $T_2$ . The external work done by the gas in this process is  $p(v_2 - v_1)$  erg., which is equivalent to  $p(v_2 - v_1)J^{-1}$  cal., where  $J$  is the mechanical equivalent of heat. The total amount of heat supplied is  $c_p$ , so that the heat  $[c_p - p(v_2 - v_1)J^{-1}]$  cal. has been spent in increasing the kinetic energy of the gas molecules. Now since the gas is an ideal one its internal energy is entirely kinetic and this depends only upon the temperature. Thus the two expressions we have obtained for the increase in kinetic energy must be identical since the range of temperature is the same in each case, *i. e.*

$$c_p - p(v_2 - v_1)J^{-1} = c_v,$$

or  $c_p - \mathcal{R}(T_2 - T_1)J^{-1} = c_v, \dots [\mathcal{R} = \text{gas constant per gramme}],$

or  $c_p - c_v = \mathcal{R}J^{-1} \quad [\because T_2 - T_1 = 1].$

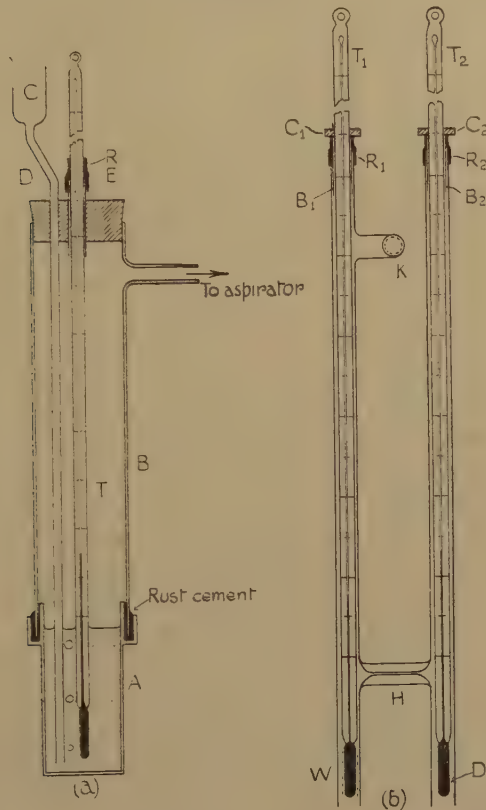
---

\* Joly, Phil. Trans. A, vol. clxxxii, p. 73 (1891).

*On a Simple Method of Constructing a Regnault's and a Wet- and Dry-bulb Hygrometer.*

"The dew-point method of determining the humidity of the atmosphere is based on straightforward physical facts and the theory is so simple that it inspires confidence in the results obtained with the instrument." Thus wrote Ezer Griffith in 1922 \*. He pointed out, however, that it is essential in the case of a Regnault hygrometer for the difference in temperature between the thermometer in the ether and

Fig. 14.



Hygrometers :—(a) Regnault's, (b) a wet- and dry-bulb hygrometer of the ventilated type.

that of the external surface of the thimble on which the dew is deposited to be small. In one form of instrument, sold for teaching purposes, a silver thimble is cemented on to the closed end of a glass tube by means of plaster of Paris, so that the above temperature difference cannot be small. Such instruments are still in use and it is the purpose of this

\* Discussion on Hygrometry, Proc. Phys. Soc. vol. xxxiv. (1922).

note to describe a method of constructing a Regnault hygrometer so that the above defect may be overcome quite easily.

A brass cap A, shaped as in fig. 14 (a), is turned out of solid brass, the walls being made as thin as possible. At the top of A there is a groove into which there fits loosely a pyrex glass tube B. A side-tube from B is connected to an aspirator when the instrument is in use. The upper end of B is provided with a rubber bung through which pass a tube D, to which a small funnel C is attached, and a short piece of brass tubing E. A thermometer T passes through E and is held in position by a short piece of rubber tubing. In this way the thermometer may be easily removed so that its corrections at the two fixed points may be determined. To effect a junction between the glass tube and the brass cap a so-called rust-cement is used. This is made by mixing together powdered ammonium chloride, flowers of sulphur, and iron filings in the proportions 1 : 2 : 80; this mixture is wetted and the annular space at the top of A filled with it. The pyrex tube is then placed in position and in a few hours the paste hardens: it is then coated with a layer of bakelite varnish to prevent any excessive formation of rust. The instrument is then used in the usual manner: the ether should be removed at once after an experiment since it penetrates slowly into the cement.

Again, in the same discussion on hygrometry by Griffiths, we read, "Whilst the unventilated type of wet- and dry-bulb hygrometer is a notoriously unreliable instrument, it does not appear sufficiently well known that by so modifying the construction that air at a definite velocity is drawn past the thermometer bulbs, it can be converted into an instrument of satisfactory precision." A suitable form of instrument for class use is shown in fig. 14 (b). The whole is constructed from glass tubing and the method of assembly needs little comment. The mercury-in-glass thermometers  $T_1$  and  $T_2$  are kept in position by being fitted into short lengths of brass tubing  $B_1$  and  $B_2$  respectively, which just slide into the glass tubes: the brass tubes are held by means of rubber connections  $R_1$  and  $R_2$ , while thick rubber bands  $C_1$  and  $C_2$  prevent the thermometers from falling through the glass tubes. This method of mounting the thermometers permits them to be withdrawn easily so that they may be standardized.

The bulb of the wet-bulb thermometer W is covered with one layer of filter paper, held in position with wire from a piece of flex, and is easily moistened with water. K is a tube leading to a water pump which provides a means of drawing air rapidly past W. A current of air is kept moving past D, the bulb of the second thermometer, by using a short capillary tube H to join the upright limbs of the instrument: if H is too wide the air flow past D must be reduced by partially closing the opening at D in any convenient manner. To calculate from the observations the relative humidity of the atmosphere under existing conditions the table given in Child's 'Physical Constants,' p. 29, is convenient.



*On the Use of an Elementary Type of Sonometer to Measure the Frequency of an a.c. Supply.*

In this experiment use is made of the formula  $f = (1/2l)\sqrt{F/\mu}$ , where  $f$  is the frequency of the fundamental mode of vibration of a wire of length  $l$ , mass per unit length  $\mu$ , when stretched by a force  $F$ . If such a wire carries an alternating current and the centre of the wire lies in a strong magnetic field, then the wire will resonate when its fundamental frequency is equal to that of the a.c. supply.

Now in order that the tension in the wire shall be very nearly equal to the stretching force, it is advisable to arrange that the wire is almost vertical—it cannot be quite vertical since true nodes would not exist permanently and continuously at the bridges. If  $M$  is the mass of the load in the scale-pan the tension in the wire will be  $F = (M+m)g$ , where  $m$  is a correction for the mass of the pan and, in effect, includes a correction for the change in tension as the wire passes over the bridge:  $g$  is the intensity of gravity. Hence  $f = (1/2l)\sqrt{(M+m)g/\mu}$ . This may be written  $(4\mu f^2/g)l^2 = M+m$ , which is equivalent to  $ax = y + b$ , so that from the slope of the straight line obtained by plotting  $x = l^2$ ,  $y = M$ ,  $f$  may be found.

To check the result obtained, the distance between the bridges may be doubled: let it be  $L$ . The wire will then resonate when its fundamental is one-half the frequency of the a.c. supply under investigation. Then  $f = (1/L)\sqrt{(M+m)g/\mu}$ , so that a similar graphical method again yields  $f$ .

A wire suitable for this experiment is made by stretching slightly a piece of copper wire [S.W.G. No. 20] about 3 metres long. A portion of this wire is used as the sonometer wire, the rest being kept as a coil of known length from which  $\mu$  may be determined.

The method gives for 50 cycles a.c. values easily correct to within 1 per cent. and, with care, the error may be reduced to less than 0.5 per cent.

*Note on a Stroboscope.*

To use this instrument satisfactorily it is necessary that the prongs of the tuning-fork whose frequency is being found should not touch the supports of the solenoid, and yet the amplitude of the prongs must be such that two glimpses of the rotating disc are seen in each complete period of the fork. If this condition is not fulfilled the pattern is fuzzy, even when it is stationary, and the clearness to be gained by increasing the amplitude of the fork is remarkable. In fact, when a clear stationary pattern has been obtained, the amplitude of the prongs should be decreased by lowering the magnitude of the exciting current, when the stationary pattern at once becomes blurred.

In the usual laboratory arrangement the tuning-fork is loaded with small metal plates having the necessary slots in them. Thus it is the frequency of a loaded fork which is found. This frequency may be 2 or more per cent. lower in value than that of the unloaded fork, so that

it is of interest to check the result for the loaded fork by an independent method. A sonometer at once suggests itself. If this is mounted horizontally on the table carrying the vibrating tuning-fork, resonance is readily observed with the naked eye, so that there is no difficulty in tuning the wire.

*On Tuning the Wire of a Vertical Sonometer.*

When a vertical sonometer is used to determine the frequency of a tuning-fork, students who are not musical find a difficulty in judging when the wire is in tune with the fork. With horizontal sonometers riders are used: they can be used even when the wire is vertical. It is only necessary to cut out a piece of paper 6 mm.  $\times$  4 mm. This is then slotted about two-thirds of the way along its major axis and can then be held readily by the wire. It is thrown off violently when the wire is in tune with the fork whose frequency is being determined.

[To be continued.]

LIII. A "*Tangent-law*" Photometer.

By the Rev. G. D. YARNOLD, University College, Nottingham\*.

[Received July 29, 1946.]

[Plate XIII.]

THE instrument described in this note may be of interest to teachers of physics in Universities and elsewhere, since it provides a novel and interesting departure from conventional methods of visual photometry, over which it possesses certain distinct advantages when the highest accuracy is not required. In the first place, an optical bench of great length is not necessary; secondly, the photometer head remains in a fixed position relative to the light sources, so facilitating the necessary adjustments; while thirdly, comparisons of extended sources can readily be made.

The photometer (which may be regarded as a variation of the wedge photometer) consists of a small white plaster cube, or right-angled prism, mounted on a turntable, whose axis of rotation lies in the edge O of the cube (figure and Pl. XIII.). The photometer head is placed so that O lies on the line joining the centres of two sources of light, which are assumed for the present to be point sources. Each of the faces OP, OQ is illuminated obliquely by light from one source only; and, by rotation of the turntable, the two illuminations are made equal at the corner. Let the

\* Communicated by Prof. L. F. Bates, Ph.D., D.Sc.







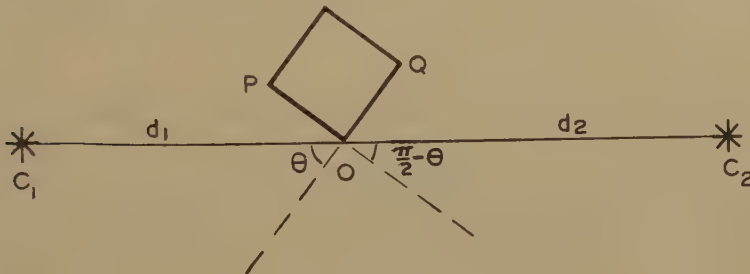
candle-powers of the sources be  $C_1$  and  $C_2$ , and let their distances from  $O$  be  $d_1$  and  $d_2$  respectively. Let the angles of incidence of light on the two faces be  $\theta$  and  $\left(\frac{\pi}{2} - \theta\right)$  respectively. Then, when both surfaces are equally illuminated,

$$\frac{C_1}{C_2} = \left(\frac{d_1}{d_2}\right)^2 \tan \theta. \quad (1)$$

Clearly, the photometer may be "balanced" whatever the values of  $d_1$  and  $d_2$ . Normally, however, when the candle-powers are of the same order of magnitude, it is desirable to make  $d_1$  and  $d_2$  equal, when the fundamental equation reduces to the more simple form,

$$\frac{C_1}{C_2} = \tan \theta. \quad (2)$$

The zero, from which the angle  $\theta$  is measured, is readily found by rotating the turntable until the face  $OQ$  is completely dark. This criterion is quite



sharp, and, with a little practice, individual settings both of this and also for equality of illumination may be made to within 1 degree. A simple test of the fundamental equation (1) may be made using motor head-lamp bulbs as sources. The filament of such bulbs is usually in the form of a short helix, which, if set parallel to the edge of the cube, is effectively a point source.

It has been assumed implicitly so far that the faces of the plaster cube are perfectly diffusing surfaces, and therefore have the same apparent brightness in all directions, a condition which is not in fact satisfied by actual surfaces. Satisfactory comparisons are possible, however, provided (a) the surfaces are viewed along the bisector of the angle of the cube, and (b) the sources are interchanged. Errors may also occur due to imperfect alignment of the sources with the edge of the cube, and to inaccuracy in the angle of the cube. With care, the former error can be made sufficiently small to have no appreciable effect. The latter error is more important, since it is not easy in practice to attain an accuracy much better than half a degree in the angle of the plaster cube.

Let us suppose that the angle of the "cube" exceeds a right angle by a small angle  $\delta$ . We now obtain

$$\frac{C_1}{C_2} = \left(\frac{d_1}{d_2}\right)^2 \cdot \tan \theta \cdot (1 - \delta \cot \theta).$$

On interchanging the lamps in their holders, we have, on re-balancing,

$$\frac{C_2}{C_1} = \left(\frac{d_1}{d_2}\right)^2 \tan \theta' \cdot (1 - \delta \cot \theta'),$$

where  $\theta'$  is the new angle of incidence of light on the face OP.

Thus, 
$$\frac{C_1}{C_2} = \sqrt{\frac{\tan \theta}{\tan \theta'}} \cdot \left[1 + \frac{\delta}{2}(\cot \theta' - \cot \theta)\right]. \quad \dots \quad (3)$$

It should be noted that equation (3) holds whatever the values of  $d_1$  and  $d_2$ . Exact equality of  $d_1$  and  $d_2$  is therefore unnecessary and, in fact, no distances need be measured, provided the lamps are accurately interchanged. The correction for inaccuracy of the angle of the cube is of little importance when the sources are of nearly equal power, but becomes significant when the ratio  $C_1/C_2$  differs considerably from unity. The error  $\delta$  may be determined with sufficient accuracy by adjusting the two faces in turn to zero illumination.

Consider now a light source consisting of a disc of radius  $a$ , of uniform intrinsic brightness  $c$  (candle-power per unit area), so placed that its axis of symmetry passes normally through the edge O of the photometer cube, at a distance  $d$  from the centre of the source. The intensity of illumination of the face of the cube, on which the angle of incidence is  $\theta$ , is given to a first approximation by

$$I = \frac{\pi a^2 c \cdot \cos \theta}{d^2} \left(1 - \frac{a^2}{d^2}\right). \quad \dots \quad (4)$$

The candle-power of the equivalent point source is thus

$$C = \pi a^2 c \left(1 - \frac{a^2}{d^2}\right), \quad \dots \quad (5)$$

and reduces to  $\pi a^2 c$  as the ratio  $a/d$  becomes vanishingly small. Since two extended sources, of equal intrinsic brightness, give equal intensities of illumination at any point from which they subtend the same solid angle, it follows that equations (4) and (5) apply also to a uniform spherical source of light, such as a frosted globe. Thus, two such extended sources, placed at equal distances from the photometer, may be compared by equation (2), or, more accurately, by equation (3), provided  $C_1$  and  $C_2$  now denote the candle-powers of the equivalent point sources.

The method may conveniently be tested by comparing three lamps. The author is indebted to Mr. F. A. Underwood, B.Sc., for the inter-comparison of three pearl lamps of 100, 60 and 40 watt rating respectively, the results of which are shown in the table.

TABLE.

Lamps :	100 watt and 60 watt.	60 watt and 40 watt.	40 watt and 100 watt.
Ratio :	2.1 <sub>0</sub>	1.6 <sub>3</sub>	0.27 <sub>6</sub>

In each case the lamps were placed at 50 cm. from the photometer. The diameters range from 6 to 8 cm. The product of the three ratios, as measured, is 0.95, instead of the theoretical value unity. In view of the fact that no precautions were taken to stabilize the currents carried by the lamps, this agreement is regarded as most satisfactory.

The author wishes to express his thanks to Mr. H. Watson, who was responsible for the construction of the photometer.

#### LIV. Generalized Routh-Hurwitz Discriminant.

(An Extension of the Theorems of Sturm\*, Routh†, and Hurwitz‡, on the Roots of Polynomial Equations.)

By SEYMOUR SHERMAN§.

[Received February 14, 1946.]

##### 1. Introduction.

A FINITE numerical algorithm is presented for finding the number of such roots of a polynomial with complex coefficients as lie in a generalized polygon or linear fractional transformation thereof. These regions include line segments (finite, semi-finite, or infinite), half-planes, infinite strips, quadrants, polygons, circles, circular arcs, and circular sectors. While this study was inspired by the needs of airplane flutter computations, where the polynomials have complex coefficients and the critical regions are the lower half-plane and the 270° sector consisting of quadrants II, III and IV, it can also be applied to statistical computations||, where the critical region is the unit circle. The method is easily adapted to airplane and helicopter dynamical stability computations, ballistical problems, the study of retarded control systems¶, computations for the stability of periodic dynamical systems, and helicopter ground resonance calculations. It is suspected that this method can be applied to other calculations arising in mechanical, aeronautical, and electrical engineering.

\* See Reference (1).

† See References (7) and (7a).

‡ First proved in Reference (4). Another proof can be found in Reference (6).

§ Communicated by the Author.

|| See Reference (8).

¶ See Reference (5).

2. *Roots on a Line Segment.*

Let  $f(z) = \sum_{\kappa=0}^n p_{\kappa} z^{n-\kappa} = 0$  be a polynomial equation with complex coefficients. We wish to consider the number of roots of  $f(z) = 0$  lying on a line segment (closed, open, half-open, finite, infinite) AB. By applying the euclidean algorithm to  $f(z)$  and  $f'(z)$  we get, in a finite number of steps,  $d(z) = (f, f')$ , the greatest common factor of  $f(z)$  and  $f'(z)$ . We see that  $f_1(z) = f/d$  is a polynomial with no multiple roots. Continuing this process, we get

$$f(z) = \prod_{\kappa=1}^r f_{\kappa}(z),$$

where each  $f_{\kappa}(z)$  has only simple roots and  $f_{\kappa+1}$  is a factor of  $f_{\kappa}$ . Our problem has, therefore, been reduced to considering the roots of  $f_{\kappa}$ . In a finite number of operations of rotating and translating axes and normalizing the equation, we reduce the problem to considering the roots of

$$f_{\kappa}(z) = \sum_{j=0}^n p_j^{\kappa} z^{n-j} = 0,$$

where AB now lies on the real axis and  $p_0^{\kappa} = 1$ .

Letting  $z = x$ , we get

$$f_{\kappa}(x) = S_1(x) + iS_2(x),$$

where degree  $S_1(x) > \text{degree } S_2(x)$  and  $S_1(x)$ ,  $S_2(x)$  have real coefficients. If  $S_2(x) \equiv 0$ , then  $f(z)$  is a polynomial with real coefficients. If  $S_2(x) \not\equiv 0$ , then consider the sequence

$$S_{k-2}(x) = q_{k-1}(x)S_{k-1}(x) - S_k(x), \quad 2 < k \leq N,$$

where  $q_k(x)$ ,  $S_k(x)$  are polynomials in  $x$ , degree  $S_{k+1} < \text{degree } S_k$ , and degree  $S_N(x) = 0$ . If  $S_N(x) \equiv 0$ , then  $S_0(x)$  and  $S_1(x)$  have no polynomial factor with real coefficients and  $f_{\kappa}(x)$  has no root on AB. If  $S_N \not\equiv 0$ , then  $f_{\kappa}(z) = S_{N-1}\phi(z)$ , where  $S_{N-1}$  is a polynomial with real coefficients and the same real roots as  $f_{\kappa}(z)$ , and  $\phi(z)$  has no real polynomial factors. By Sturm's theorem, we calculate the number of roots of  $S_{N-1} = 0$  lying on AB, and hence, the number of roots of  $f_{\kappa}(z) = 0$  lying on AB. Thus

*Theorem I.*—Given a polynomial equation

$$f(z) = \sum_{\kappa=0}^n p_{\kappa} z^{n-\kappa} = 0,$$

with complex coefficients and a line segment (closed, open, half-open, finite, infinite) one can, in a finite number of operations, determine both the number of roots lying on the line segment and their multiplicity.

3. *Cauchy's Index Theorem.*

It is a well-known theorem of Cauchy and a simple consequence of the theorem of ref. <sup>(10)</sup>, pp. 115–16, on the poles and zeros of meromorphic functions, that

If  $w = f(z)$  is a polynomial function of  $z$  and  $z$  describes a simple closed



contour in the  $z$ -plane, where the path of  $z$  intersects none of the roots of  $w=0$ , then the number of roots within the  $z$ -contour is equal to the number of times the  $w$ -contour encircles the origin.

Let

$$w(z)=P+iQ,$$

where both  $P$  and  $Q$  are real and for no value of  $z$  are both  $P$  and  $Q$  equal to zero. Let  $\alpha$  be the number of times that  $P/Q$  passes through zero from  $+$  to  $-$  and  $\beta$  be the number of times that  $P/Q$  passes through zero from  $-$  to  $+$  as  $z$  travels in a counter-clockwise direction from the  $z$ -contour. Then  $\frac{1}{2}(\alpha-\beta)$  is equal to the number of times the  $w$ -contour encircles the origin, and so, by Cauchy's Theorem, this is equal to the number of roots within the  $z$ -contour. In computing  $\alpha$  and  $\beta$ , we may notice that  $P$  may be zero for a whole interval of values of  $z$ . For this case, the  $\alpha$  and  $\beta$  are computed as before with no multiplicity accruing from the fact that  $P=0$  for an interval of  $z$ 's.

#### 4. Infinite Sector Theorem.

We wish to consider the number of roots of  $w=f(z)=\sum_{\kappa=0}^n p_{\kappa}z^{n-\kappa}=0$  lying in an infinite sector of vertex angle  $\Theta_2-\Theta_1>0$ . From the argument of Section 2, we see that this reduces to considering the roots of  $f(z)=0$ , where  $f(z)$  has no multiple roots. After a finite number of operations of rotating and translating axes, normalizing the equation and factoring out  $z$ , we see that the problem is equivalent to finding how many roots of

$$f(z)=\sum_{\kappa=0}^n p_{\kappa}z^{n-\kappa}=0,$$

where  $p_0=1$  and  $p_n\neq 0$  lie in an infinite sector with vertex at the origin and going counter-clockwise from the ray  $\Theta=\Theta_1\geq 0$  to the ray  $\Theta=\Theta_2\geq 0$ ,  $\cot \kappa\Theta_1\neq 0\neq \cot \kappa\Theta_2$ ,  $1\leq \kappa\leq n$ .

We now follow Routh somewhat, but have a more general  $z$ -contour (an infinite sector instead of an infinite half-plane) and a more general equation (with complex coefficients instead of real coefficients). The theorem of Cauchy (Section 3) is used where the  $z$ -contour is a sector  $(\Theta_1, \Theta_2)$  with large radius  $R$ . Let  $z$  vary along the vector  $\Theta=\Theta_1$ ,

$$z=re^{i\Theta_1}, \quad r\geq 0.$$

Writing  $p_{\kappa}=R_{\kappa}e^{i\alpha_{\kappa}}$ ,  $0\leq \kappa\leq n$ ,  $R_{\kappa}$  real,  $\alpha_{\kappa}$  real,

$$\begin{aligned} \text{we get} \quad f(z) &= \sum_{\kappa=0}^n r^{n-\kappa} [R_{\kappa} \cos\{(n-\kappa)\Theta_1+\alpha_{\kappa}\}] \\ &\quad + i \sum_{\kappa=0}^n r^{n-\kappa} [R_{\kappa} \sin\{(n-\kappa)\Theta_1+\alpha_{\kappa}\}] \\ &= S_1(\Theta_1, r) + iS_2(\Theta_1, r), \end{aligned}$$

where  $S_1$  and  $S_2$  are real polynomials in  $r$ . As in Section 2, we construct the Sturm sequence  $S_1, S_2, \dots, S_{N_1}$ . If  $S_{N_1}\equiv 0$ , then  $f(r)$  has a real polynomial factor and  $f(z)$  has a factor which may reduce essentially to the

trivial one  $f(z)$  itself, when  $N=2$ . In any event, by operating in this manner for the rays  $\Theta_1$  and  $\Theta_2$ , we factor

$$f(z) = \prod_{\kappa=0}^t g_{\kappa}(z)$$

into a constant factor  $g_0$  and a set of polynomials  $g_{\kappa}(z)$ ,  $1 \leq \kappa \leq t$  where either (1)  $g_{\kappa}$  has a complete Sturm sequence  $S_{N_1}(\Theta_1) \equiv \equiv 0 \equiv S_{N_1}(\Theta_2)$  for both  $\Theta_1$  and  $\Theta_2$ , (2)  $g_{\kappa}$  is real for both  $\Theta_1$  and  $\Theta_2$ , or (3)  $g_{\kappa}$  is real on one of the rays but has a complete Sturm sequence on the other ray.

Case 1 occurs most often in practice. In this case no roots of  $g(z)=0$  lie on the boundary of the sector or its extensions. Here we follow Routh in counting  $\frac{1}{2}(\alpha-\beta)$  for the sector whose radius becomes infinite. As  $r$  varies from 0 to  $R$ , with  $\Theta_1$  fixed, one of the functions  $S_{\kappa}(\Theta_1)$ ,  $1 < \kappa < N_1$  may vanish. In that case, the classical argument of Sturm's Theorem shows that

$$V(\Theta_1, r) = V\{S_1(\Theta_1, r), S_2(\Theta_1, r), \dots, S_{N_1}(\Theta_1, r)\},$$

the number of variations in sign in the Sturm sequence, does not change as  $r$  goes through a zero of  $S_{\kappa}$ ,  $1 < \kappa < N_1$ . As  $r$  goes from 0 to  $R$ ,  $V(\Theta_1, r)$  can change only because of a change in sign in one of the functions  $S_1, S_{N_1}$ . Since  $S_{N_1}$  is a constant, we do not get any changes in  $V$  due to that function; therefore,  $V$  can change only because the leading term  $S_1$  goes through a zero. Because our hypothesis that  $w(z) \neq 0$  on the boundary, we see that  $S_1(\Theta_1, r)$  and  $S_2(\Theta_1, r)$  may not become zero at the same point along the ray  $\Theta_1$ . Therefore, as the ratio  $S_1/S_2$  passes through zero from  $+$  to  $-$ , the value of  $V(\Theta_1, r)$  increases by one, and, when  $S_1/S_2$  passes through zero from  $-$  to  $+$ ,  $V(\Theta_1, r)$  decreases by one. As we go counter-clockwise from  $R$  to 0 on the  $\Theta_2$  vector and from 0 to  $R$  on the  $\Theta_1$  vector, the contribution to  $\frac{1}{2}(\alpha-\beta)$  is

$$\frac{1}{2}\{-V(\Theta_2, R) + V(\Theta_2, 0) - V(\Theta_1, 0) + V(\Theta_1, R)\}.$$

We now wish to consider the contribution to  $\frac{1}{2}(\alpha-\beta)$  on the arc of a circle of radius  $R$  and going from  $\Theta_1$  to  $\Theta_2$ . As  $R$  becomes sufficiently large, both the real and imaginary parts of  $f(z)$  are dominated by the real and imaginary parts of  $z^n$ . But, as  $z$  goes along the circular arc,  $(\alpha-\beta)$

equals the number of times  $z^n$  crosses the ray  $\Theta = \frac{\Pi}{2}$  and the ray  $\Theta = \frac{3\Pi}{2}$ ,

and so  $\alpha-\beta$  is equal to the number of odd integers lying between  $\frac{2n\Theta_1}{\Pi}$  and  $\frac{2n\Theta_2}{\Pi}$ . Let us call this quantity  $\theta(n, \Theta_1, \Theta_2)$ . For sufficiently great  $R$ ,

$$\frac{\alpha-\beta}{2} = \frac{1}{2}\{-V(\Theta_2, R) + V(\Theta_2, 0) - V(\Theta_1, 0) + V(\Theta_1, R) + \theta(n, \Theta_1, \Theta_2)\}.$$

But

$$V\{S_1(\Theta_1, r), S_2(\Theta_1, r), \dots, S_{N_1}(\Theta_1, r)\} \xrightarrow{r \rightarrow \infty} \infty$$

$$V\{p_0', p_0^2, \dots, p_0^{N_1}\} = V(\Theta_1, \infty),$$

where  $p_0', p_0^2, \dots, p_0^{N_1}$ , are the leading coefficients in  $S_1, \dots, S_{N_1}$ , and, of course,  $V(\Theta_1, 0)$  is the number of variations in sign of the constant terms in the functions  $S_1, S_2, \dots, S_{N_1}$ . Hence, as we go counter-clockwise around the sector

$$\frac{\alpha - \beta}{2} = \frac{1}{2} \{ -V(\Theta_2, \infty) + V(\Theta_2, 0) - V(\Theta_1, 0) + V(\Theta_1, \infty) + \theta(n, \Theta_1, \Theta_2) \}.$$

The same expression is pertinent to cases (2) and (3) with the simplification that in the cases where  $g_\kappa$  is real on the ray, because  $w \neq 0$  on the ray, the contribution to  $\frac{\alpha - \beta}{2}$  along the ray is zero. We therefore have proved

*Theorem II.*—Given an infinite sector in the  $z$ -plane and a polynomial  $w(z)$ , which does not vanish on the boundary, we can, in a finite number of steps, determine the number of roots of  $w=0$ , lying in the interior of the sector.

This method can be modified to take care of the case in which  $w(z)$  has some roots on the boundary of the sector. By the method of Theorem I, one can determine the number of roots on the boundary. For sufficiently small  $\epsilon > 0$  there will be no roots on the rays  $\Theta_1 - \epsilon$ ,  $\Theta_1 + \epsilon$ ,  $\Theta_2 - \epsilon$ ,  $\Theta_2 + \epsilon$ . By the methods of Theorem II, one can find the number of roots interior to the sector bounded by the rays  $\Theta_1 + \epsilon$  and  $\Theta_2 - \epsilon$  and also the number of roots interior to the sector bounded by the rays  $\Theta_1 - \epsilon$  and  $\Theta_2 + \epsilon$ . If the number of roots in these two sectors differs by the number of roots previously found to be on the boundary, then the number of roots in the inner sector is equal to the number of roots in the interior of the original sector. This procedure, however, is not necessarily finite.

### 5. Generalized Polygon Theorem.

Any convex polygon (possibly infinite) has as its complement a finite number of infinite open sectors plus a finite number of line segments. Thus, one can determine the number of roots of  $w(z)=0$  lying in a polygon of the  $z$ -plane and the number lying on the boundary of the polygon. Since any generalized plane polygon, *i. e.* a plane area bounded by a finite number of line segments, can be broken up into a finite number of possibly infinite convex polygons, we have shown

*Theorem III.*—Given a generalized plane polygon  $A$ , and a polynomial equation  $w(z)=0$ , one can, in a finite number of steps, determine the number of such roots of  $w(z)=0$  as lie in  $A$ .

### 6. Linear transformations of generalized polygons.

Let us consider  $A$ , a generalized polygon in the  $z$ -plane. If we now apply the transformation  $z'=1/z$ ,  $A$  is mapped into  $A'$ . The zeros

$$\{z_1', z_2', \dots, z_n'\} \text{ of } f(z') = \sum_{\kappa=0}^n a_\kappa (z')^{n-\kappa} = 0$$

correspond to

$$\left\{ z_1 = \frac{1}{z_1'}, z_2 = \frac{1}{z_2'}, \dots, z_n = \frac{1}{z_n'} \right\},$$

the zeros of 
$$g(z)/z^n = \sum_{\kappa=0}^n a_{\kappa} \left( \frac{1}{z} \right)^{n-\kappa} = 0.$$

Thus, if we determine the number of zeros of  $g(z)=0$  in  $A$ , we have the number of zeros of  $f(z')=0$  in  $A'$ . By this means we can find the number of such roots of a given polynomial as lie in the inverse of a generalized polygon. Given a polynomial  $w(z)$  with roots  $\{z_1, \dots, z_n\}$  we can, in a finite number of steps, construct a polynomial  $v(z)$  of the same degree with roots  $\{z_1'=\alpha z_1+\beta, \dots, z_n'=\alpha z_n+\beta\}$ . Thus, by combining the two previous statements with Theorem III, we see that given a generalized polygon  $A$ , we can find the number of such roots of a polynomial equation  $w(z)=0$  as lie in any linear fractional transform of  $A$ . We can, therefore, find the number of such roots of a polynomial as lie in a circle or sector of a circle.

### 7. Applications.

The problem of computing the flutter velocity of an airplane lift or control surface reduces at one point to the question of whether or not all the roots of a complex polynomial equation lie in a given 270-degree sector (quadrants II, III and IV of the complex plane). The problem may be treated in two stages applying (1) a sufficient condition for stability, which is that all of the roots lie in the lower half-plane, and (2) a necessary and sufficient condition for stability, which is that all of the roots lie in the aforementioned 270-degree sector.

Suppose we have a polynomial

$$\begin{aligned} f(z) &= \sum_{\kappa=0}^n (p_{\kappa}^0 + i q_{\kappa}^0) z^{n-\kappa} \\ &= S_0(x) + i S_1(x), \end{aligned}$$

which does not vanish on the real axis and where  $p_{\kappa}^0, q_{\kappa}^0, 0 \leq \kappa \leq n$ ,  $S_0$ , and  $S_1$  are real and  $p_0^0 + i q_0^0 = 1$ . Then we apply the infinite sector theorem to the case  $\Theta_1=0$  and  $\Theta_2=\Pi$  in order to secure a condition on the coefficients so that no roots lie in the upper half-plane.

We get

$$\theta(n, 0, \Pi) = n$$

$$\frac{\alpha - \beta}{2} = 0$$

and 
$$\begin{aligned} 0 &= \{-V(\Pi, \infty) + V(\Pi, 0) - V(0, 0) + V(0, \infty) + n\} \\ &= \{-V(\Pi, \infty) + V(0, \infty) + n\}. \end{aligned}$$

Since  $V(0, \infty)$  and  $V(\Pi, \infty)$  are positive numbers, each not greater than  $n$ , then, in order that the above equation holds, we must have  $V(0, \infty)=0$  and  $V(\Pi, \infty)=n$ , which are, fortunately, equivalent conditions on  $V$ . We now set up the Sturm sequence

$$\{S_0, S_1, \dots, S_n\}.$$



where

$$S_j = \sum_{\kappa=0}^n p_{\kappa}^j x^{n-j-\kappa},$$

and our requirement on  $V$  yields  $p_0^j > 0$ ,  $0 \leq j \leq n$ . If we multiply each of the  $p_0^j$  by a positive number, the sign of the terms will be unaltered.

We now follow Frazer, Duncan and Collar†, *Elementary Matrices* <sup>(9)</sup>.

If 
$$S_0 = \sum_{\kappa=0}^n p_{\kappa}^0 x^{n-\kappa},$$

then let  $T_0$  be the row:  $a_0^0, a_1^0, \dots$ ,

where

$$\begin{aligned} a_{\kappa}^0 &= p_{\kappa}^0, \quad \kappa \leq n \\ a_{\kappa}^0 &= 0, \quad \kappa > n. \end{aligned}$$

Moreover, if

$$S_1 = \sum_{\kappa=0}^{n-1} p_{\kappa}^1 x^{n-1-\kappa},$$

then let  $T_1$  be the row:

$$a_0', a_1', \dots,$$

where

$$\begin{aligned} a_{\kappa}' &= p_{\kappa}', \quad \kappa \leq n-1 \\ a_{\kappa}' &= 0, \quad \kappa > n-1. \end{aligned}$$

We now form the sequence  $T_0, T_1, T_1^*, \widetilde{T}_2, T_2^*, \dots, \widetilde{T}_n$  of rows. The elements of  $T_1^*$  are given by a cross-multiplication process

$$a_{\kappa}'^* = a_0' a_{\kappa+1}^0 - a_{\kappa+1}' a_0^0, \quad 0 \leq \kappa.$$

The elements of  $\widetilde{T}_2$  are derived by operating in the same way with rows  $T_0$  and  $T_1^*$ . Thus,

$$\tilde{a}_{\kappa}^2 = a_0'^* a_{\kappa+1}^0 - a_{\kappa+1}'^* a_0^0, \quad 0 \leq \kappa.$$

In general, the elements of  $T_p^*$ ,  $p > 1$ , are given by

$$a_{\kappa}^{p*} = \tilde{a}_0^p \tilde{a}_{\kappa+1}^{p-1} - \tilde{a}_{\kappa+1}^p \tilde{a}_0^{p-1}, \quad 0 \leq \kappa.$$

while the elements of  $\widetilde{T}_p$ ,  $p > 1$ , are given by

$$\tilde{a}_{\kappa}^p = a_0^{p-1*} \tilde{a}_{\kappa+1}^{p-1} - a_{\kappa+1}^{p-1*} \tilde{a}_0^{p-1}, \quad 0 \leq \kappa.$$

A necessary and sufficient condition that the leading elements of  $T_0, T_1, \widetilde{T}_2, \dots, \widetilde{T}_n$  be each positive is that each of the leading coefficients of  $S_0, S_1, S_2, \dots, S_n$  be positive.

As an example of the applications of this procedure, let us take from reference <sup>(9)</sup> the actual flutter equation

$$f(z) = z^3 - (-4.3306 + 0.5400i)z^2 + (5.6056 - 1.3747i)z - 1.9857 + 0.6531i = 0,$$

where we assume that the coefficients are given exactly. The sequence

---

† See Reference <sup>(2)</sup>.

$T_0, T_1, T_1^*, \widetilde{T}_2, T_2^*, \widetilde{T}_3$  is given in the following table :—

$T_0$	1	-4.3306	5.6056	-1.9857
$T_1$	0.5400	-1.3747	0.6531	0
$T_1^*$	-0.9638	2.3739	-1.0723	0
$\widetilde{T}_3$	0.0430	-0.0504	0	0
$T_2^*$	-0.0319	0.0281	0	0
$\widetilde{T}_3$	0.0004	0	0	0

In this case the leading elements of  $T_0, T_1, \widetilde{T}_2, \widetilde{T}_3$  are all positive, so none of the roots lies in the upper half-plane and the sufficient condition for stability is satisfied.

While the approximate solution of a flutter cubic by Newton's Method takes a skilled computer, operating a Marchant, about two and one-half hours, the same computer can, using a Marchant, apply the sufficient condition in fifteen minutes. This method, generalized to the necessary and sufficient condition for the absence of flutter, can be found in ref. <sup>(9)</sup> with an example worked out. When one of the roots is close to the real axis, the method also provides a close approximation to that root.

### 8. Half-planes.

In the more conventional instability problem, one usually has an equation

$$f(z) = \sum_{\kappa=0}^n (p_{\kappa}^0) z^{n-\kappa} = 0,$$

and a condition is sought on the coefficients so that none of the roots lies in the right half-plane. This suggests the general problem: given

$$f(z) = \sum_{\kappa=0}^n (p_{\kappa}^0 + iq_{\kappa}^0) z^{n-\kappa} = 0,$$

with no roots on the rays  $\Theta$ , and  $\Theta + \Pi$ , find a necessary and sufficient condition on the coefficients so that none of the roots lies in the open infinite sector going counter-clockwise from  $\Theta$  to  $\Theta + \Pi$ . This may be converted to the problem discussed by introducing the variable  $z' = ze^{-i\Theta}$ , and letting

$$\begin{aligned} g(z) &= e^{-in\Theta} \sum_{\kappa=0}^n (p_{\kappa}^0 + iq_{\kappa}^0) (z'e^{i\Theta})^{n-\kappa} \\ &= \sum_{\kappa=0}^n (p_{\kappa}^0 + iq_{\kappa}^0) (z')^{n-\kappa} e^{-i\kappa\Theta}; \end{aligned}$$

$g(z)$  has no roots  $z'$  in the upper half-plane if, and only if,  $f(z)$  has no roots in the sector going counter-clockwise from  $\Theta$  to  $\Theta + \Pi$ . We can deal with

$g(z)$  by means of the method given in the last section. For the usual stability problem

$$\Theta = \frac{3\Pi}{2} \text{ and } \Theta + \Pi = \frac{5\Pi}{2} \text{ and}$$

$$g(z) = \sum_{\kappa=0}^n (p_{\kappa}^0)(z')^{n-\kappa} e^{-i\kappa 3\Pi/2}.$$

### 9. Generalized Hurwitz Discriminant.

A. Hurwitz, whose attention to the standard stability problem had been aroused by A. Stodola, the famous steam turbine engineer, and some remarks in Thomson and Tait, 'Natural Philosophy,' showed by means of the theory of quadratic forms that the conditions on the coefficients of a real polynomial, in order that none of the roots has real parts, could be reduced to the requirement that a set of determinants be positive. His requirements, while not more simple to calculate than the iterative sequence of inequalities devised by Routh, nevertheless have a certain algebraic simplicity so that, while in the previous sections we have been generalizing Routh's results, we now concern ourselves with generalizing Hurwitz's result. The Hurwitz theorem may be stated:

The roots of  $\sum_{\kappa=0}^n (p_{\kappa})z^{n-\kappa}=0$ , with  $p_0>0$  and  $p_{\kappa}$  real,  $0<\kappa$ , all have negative real parts if, and only if, the determinants of the principal  $\kappa \times \kappa$ ,  $\kappa=1, \dots, n$ , sub-matrices of

$$\left\| \begin{array}{cccccc} p_1 & p_0 & 0 & 0 & 0 & \dots & 0 \\ p_3 & p_2 & p_1 & p_0 & 0 & \dots & 0 \\ p_5 & p_4 & p_3 & p_2 & p_1 & \dots & 0 \\ . & . & . & . & . & . & 0 \\ p_{2n-1} & p_{2n-2} & . & . & . & . & p_n \end{array} \right\|$$

where  $p_{\kappa}=0$ ,  $\kappa>n$ , are all positive.

We now wish to derive a result similar to that of Hurwitz but applying to the sector between  $\Theta$  and  $\Theta + \Pi$  and to a polynomial with complex coefficients. From the reasoning given earlier in this section, we see that this can be reduced to a sequence of inequalities not too unlike the sequence that Routh derived for the real case. This suggests that it might prove helpful to follow the reasoning of Frazer and Duncan<sup>(10)</sup>, who, unaware of the Hurwitz paper, had derived Hurwitz's result from the Routh theorem.

In the Routh iterative process we start with two sequences:

$$T_0: \{a_0^0, a_1^0, \dots\} \text{ and}$$

$$T_1: \{a_0', a_1', \dots\},$$

---

\* See Reference (8).

and let  $T_{n+1} = df \alpha(T_{n-1}, T_n) = \{a_0^{n+1}, a_1^{n+1}, \dots\}$ ,  $n \geq 1$ ,

where  $a_{\kappa}^{n+1} = \begin{vmatrix} a_0^n & a_0^{n-1} & 0 \\ a_1^n & a_1^{n-1} & a_0^n \\ a_{\kappa+2}^n & a_{\kappa+2}^{n-1} & a_{\kappa+1}^n \end{vmatrix}$ ,

and thus we have an infinite set  $[T_0, T_1, T_2, \dots]$  of sequences. If we choose the  $a_j^0$  and  $a_j'$ ,  $0 \leq j$ , as in Section 7, the  $T_2, T_3, T_4, \dots$  described here are the same as the  $\widetilde{T}_2, \widetilde{T}_3, \widetilde{T}_4, \dots$  described there.

We now define an infinite class of functions  $[\overline{T}_0, \overline{T}_1, \overline{T}_2, \dots]$  from ordered pairs of sequences to sequences and in terms of these functions we get a collection of sequences  $[P_j/j=0, 1, 2, \dots]$ .

Let  $P_0 = T_0 = \{a_0^0, a_1^0, a_2^0, \dots\}$

and  $P_1 = T_1 = \{a_0', a_1', \dots\}$ .

For sequences A, B, we define

$$\overline{T}_2(A, B) = \alpha(A, B)$$

$$\overline{T}_n(A, B) = df \overline{T}_{n-1}(B, \alpha(A, B)), \quad n \geq 3.$$

We define  $P_n = df \overline{T}_n(T_0, T_1)$ ,  $n \geq 2$ ,

and note that  $P_2 = \overline{T}_2(T_0, T_1) = \alpha(T_0, T_1) = T_2$ ,

$$P_3 = \overline{T}_3(P_0, P_1) = \alpha(P_1, P_2) = T_3.$$

$$P_n = \overline{T}_n(P_0, P_1) = \overline{T}_{n-1}(P_1, P_2), \quad n \geq 3.$$

We wish to prove that

$$\star \quad P_n = T_n, \quad n = 1, 2, 3, \dots$$

We note that the equality holds for  $n = 1, 2, 3$ , and assume that it holds for  $n \leq N-1$ , where  $N \geq 3$ . Then

$$\begin{aligned} P_N &= \overline{T}_N(P_0, P_1) = \overline{T}_{N-1}(P_1, P_2) \\ &= \overline{T}_{N-2}(P_2, \alpha(P_1, P_2)) \\ &= \overline{T}_{N-2}(P_2, P_3) \\ &= \overline{T}_{N-3}(P_3, P_4) \\ &= \overline{T}_{N-(N-2)}(P_{N-2}, P_{N-1}) = \alpha(P_{N-2}, P_{N-1}) \\ &= \alpha(T_{N-2}, T_{N-1}) = T_N. \end{aligned}$$

Hence equation ( $\star$ ) has been proved.

We know that

$$a_0' = a_0'$$

$$a_0^2 = \begin{vmatrix} a_0' & a_0^0 & 0 \\ a_1' & a_1^0 & a_0' \\ a_2' & a_2^0 & a_1' \end{vmatrix}$$





=

$$(a_0')^{(N-2)(N-3)}(-a_0')^{N-2}$$

$$\left| \begin{array}{cccc|cc|c} a_0' & a_0^0 & 0 & 0 & 0 & & \\ a_1' & a_1^0 & a_0' & 0 & 0 & & \\ a_2' & a_2^0 & a_1' & a_0' & \left| \begin{array}{cc} a_0' & a_0^0 \\ a_1' & a_1^0 \end{array} \right| & & 0 \\ a_3' & a_3^0 & a_2' & a_1' & \left| \begin{array}{cc} a_0' & a_0^0 \\ a_2' & a_2^0 \end{array} \right| & & a_0' \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{2N-2}' & a_{2N-2}^0 & a_{2N-3}' & a_{2N-4}' & \left| \begin{array}{cc} a_0' & a_0^0 \\ a_{2N-3}' & a_{2N-3}^0 \end{array} \right| & a_{2N-5}' \left| \begin{array}{cc} a_0' & a_0^0 \\ a_N' & a_N^0 \end{array} \right| & \end{array} \right|$$

If we let the new  $\overline{\text{col } 5} = a_0^0 \overline{\text{col } 3} + \overline{\text{col } 5}$  and the new  $\overline{\text{col } (2k+1)} = a_0^0 \overline{\text{col } (2k-2)} + \overline{\text{col } (2k+1)}$ ,  $k=3, \dots, N-1$ , and simplify, we get

$$a_0^N = (a_0')^{(N-2)(N-3)}(+a_0')^{2(N-2)}(-1)^{(N-2)}$$

$$\left| \begin{array}{cccccccc} a_0' & a_0^0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ a_1' & a_1^0 & a_0' & 0 & a_0^0 & 0 & 0 & \cdot & \cdot \\ a_2' & a_2^0 & a_1' & a_0' & a_1^0 & 0 & a_0^0 & \cdot & \cdot \\ a_3' & a_3^0 & a_2' & a_1' & a_2^0 & a_0' & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{2N-2}' & a_{2N-2}^0 & a_{2N-3}' & a_{2N-4}' & a_{2N-3}^0 & \cdot & \cdot & \dots & a_N^0 \end{array} \right|$$

and by re-arranging columns, we get

$$a_0^N = (a_0')^{(N-2)(N-1)}$$

$$\left| \begin{array}{cccccccc} a_0' & a_0^0 & 0 & 0 & 0 & 0 & \cdot & \dots & 0 \\ a_1' & a_1^0 & a_0' & a_0^0 & 0 & 0 & \cdot & \cdot & \cdot \\ a_2' & a_2^0 & a_1' & a_1^0 & a_0' & a_0^0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{2N-2}' & a_{2N-2}^0 & a_{2N-3}' & a_{2N-3}^0 & \cdot & \cdot & \cdot & \dots & a_N^0 \end{array} \right|$$

It is well now to collect the result of this section and to formulate it as a theorem.

*Theorem IV.*—Generalized Hurwitz Discriminant. The complex polynomial equation

$$f(z) = \sum_{\kappa=0}^n b_{\kappa} z^{n-\kappa} = 0,$$

with non-zero leading coefficient, has no roots in the closed half-plane swept out by a semi-infinite half-ray from the origin rotating in a counter-clockwise direction from  $\Theta$  to  $\Theta + \Pi$  if, and only if, the determinants of the principal  $(2k-1) \times (2k-1)$  matrices,  $1 \leq k \leq n$ , of

$$\begin{vmatrix} a_0' & a_0^0 & 0 & 0 & 0 & 0 & . & . & 0 \\ a_1' & a_1^0 & a_0' & a_0^0 & 0 & 0 & . & . & . \\ a_2' & a_2^0 & a_1' & a_1^0 & a_0' & a_0^0 & . & . & . \\ . & . & . & . & . & . & . & . & . \\ a_{2n-2}' & a_{2n-2}^0 & a_{2n-3}' & a_{2n-3}^0 & . & . & . & . & a_n^0 \end{vmatrix}$$

are all positive, where

$$\begin{aligned} & \sum_{\kappa=0}^n a_{\kappa}^0 (z')^{n-\kappa} + i \sum_{\kappa=0}^{n-1} a_{\kappa}' (z')^{n-1-\kappa} \\ &= \frac{\sum_{\kappa=0}^n b_{\kappa} (z')^{n-\kappa} e^{i\kappa\Theta}}{b_0} \end{aligned}$$

$$a_{\kappa}^j, \text{ real, } 0 \leq k, j=0, 1,$$

$$\text{and } a_{\kappa}^j = 0, k+j > n.$$

The generalized Hurwitz discriminant is not as convenient to apply as the method (which we can call the generalized Routh discriminant) described in section 7. If we apply the generalized discriminant to the case considered in the Hurwitz Theorem (real coefficients), the requirement that all the roots lie in the open left half-plane now becomes: the determinants of the principal  $(2k-1) \times (2k-1)$  matrices,  $1 \leq k \leq n$ , of

$$\begin{vmatrix} p_1 & p_0 & 0 & 0 & 0 & 0 & . & . & 0 \\ 0 & 0 & p_1 & p_0 & 0 & 0 & . & . & . \\ -p_2 & -p_3 & 0 & 0 & p_1 & p_0 & . & . & . \\ 0 & 0 & -p_2 & -p_3 & 0 & 0 & . & . & . \\ p_4 & p_5 & 0 & 0 & -p_2 & -p_3 & . & . & . \\ . & . & . & . & . & . & . & . & \epsilon(N)p_n \end{vmatrix}$$

are all positive where  $\epsilon(N)=1$  for  $N \equiv 0$  or  $1 \pmod{4}$  and  $\epsilon(N)=-1$  otherwise. An elementary calculation reveals that the set of quotients of

successive determinants mentioned above give exactly the determinants of the Hurwitz Theorem, so that the generalized Hurwitz Theorem is consistent with it.

### 10. Appendix.

After this paper was written, the American Mathematical Society (Gibbs' Lecture on "The Control of an Elastic Fluid," by H. Bateman (Bull. Amer. Math. Soc. li. pp. 601-46 (1945), appeared. In it are various references to stability problems, the contents on pp. 607-18 being related to the contents in this paper. After discussing the Routh-Hurwitz discriminant in the real case, Bateman remarks on p. 612, "The study of equations with complex coefficients is also useful as there are some cases in which the conditions of Routh and Hurwitz are not the simplest possible conditions for pseudo-negative roots." He then discusses two dynamical problems (H. Lamb, "On Kinetic Stability," Proc. Roy. Soc. London, Ser. A, lxxx. pp. 168-77 (1908), and H. Bateman, "Stability of the Parachute and Helicopter," National Advisory Committee for Aeronautics, Report No. 80 (1920)). Their solutions depend on all of the zeros of a complex quadratic and a complex cubic polynomial, respectively, lying in the left half-plane. On pp. 613-16, conditions on the coefficients are derived in order that this be so. Since these are special cases solved here for any degree  $n$ , they could have been conveniently treated by either the method of Section 7 (generalized Routh discriminant) coupled with a rotation of the co-ordinate system, or the method of Section 9 (generalized Hurwitz discriminant).

### References.

- (1) Dickson L. E., 'First Course in the Theory of Equations,' pp. 75. (John Wiley & Sons, Inc., New York, 1922).
- (2) Frazer, R. A., Duncan, W. J., and Collar, A. R., 'Elementary Matrices and some Applications to Dynamics and Differential Equations,' pp. 151-5. (Cambridge Univ. Press, 1938.)
- (3) Frazer, R. A., and Duncan, W. J., "On the Criteria for the Stability of Small Motions," Proc. Roy. Soc. A, cxxiv. p. 642 *et seq.* (1929).
- (4) Hurwitz, A., "Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit Negativen Reellen Theilen Besitzt," *Math. Ann.* xlv. p. 273 *et seq.* (1895).
- (5) Minorsky, N., "Self-excited Oscillations in Dynamical Systems possessing Retarded Control Actions," Journal of Applied Mechanics (June 1942).
- (6) Riemann-Weber, 'Partiellen Differential Gleichungen der Mechanik und Physik,' pp. 162 *et seq.* (Mary S. Rosenberg, New York, 1943.)

---

\* In Reference (8), an elegant formulation can be found of the condition that the roots of a polynomial equation be less than one in absolute value.

† See Reference (5).

‡ In Reference (9), tabular arrangements based on References (6), (7), and (7a), and some of the developments of this paper, are given for an important stage of flutter computations and numerical examples are presented.



- (7) Routh, E. J., 'Dynamics of a System of Rigid Bodies,' 6th edn. Pt. II. pp. 221 *et seq.* (MacMillan and Co., London, 1905.)
- (7a) Routh, E. J., 'A Treatise on the Stability of a given State of Motion,' Chap. 2 and 3. (London, 1877.)
- (8) Samuelson, P. A., "Conditions that the Roots of a Polynomial be less than Unity in Absolute Value," *Ann. Math. Statist.*, xii. pp. 360-4, 1941.
- (9) Sherman, S., DiPaola, J., Frissel, H. F., 'Routh's Discriminant and Airfoil Flutter,' Curtiss Research Laboratory Report, V-241-S-1. (Buffalo, 1943.)
- (10) Titchmarsh, E. C., 'Theory of Functions, 2nd edn. (Oxford University Press, 1939.)

University of Chicago,  
Chicago, Illinois,  
U.S.A.

---

---

LV. *A Critical Examination of Milne's Kinematical Relativity.*

By WILLIAM BAND, Department of Physics, Yenching University,  
Peiping, China\*.

[Received October 22, 1945.]

*Outline of Milne's Theory*<sup>(1)</sup>.

1.1. World geometry is set up by Milne as a system of relations between a specially selected set of equivalent observers each provided with nothing more than a clock, a theodolite and some unspecified mechanism whereby they can exchange messages with each other. This mechanism must be such that it permits synchronization of the whole collection of clocks throughout all time and space, and also permits the assignment by any arbitrarily first chosen observer of a unique value to the distance between any two other observers in terms of time measures. The minimum requirement for all this is that the messages must be transmitted with a given velocity with respect to every observer everywhere and all the time, and is identified by Milne, partly for this reason, with light or wireless signals. Milne claims that this shows that the constancy of the velocity of light is an *a posteriori* consequence of the equivalence of the observers; at least it means that if the velocity of light were not in fact constant, Milne's plans for setting up world geometry could not succeed.

1.2. Milne proposes the following principle of selection: two observers are accepted as equivalent if the totality of observations A can make on B can be described by A in the same way as B can describe the totality of observations B can make on A. Then, extending this to a world-system of observers, they are said to constitute an equivalent or fundamental system if every pair is an equivalent pair in the above sense.

---

\* Communicated by the Author.

1.3. Milne applies the principle of relativity in what he claims to be a much weaker form than that used in the Einstein theory : namely, that the laws of nature are invariant in form with respect to transformations between the co-ordinates used by the observers in the fundamental system of equivalent observers.

1.4. Milne next attempts to define mathematically a system of relatively moving observers that shall satisfy his principle of selection, using the Newtonian transformation of velocities, and shows that if and only if

$$\mathbf{v} = \mathbf{r}/t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

then the system of observers with velocities  $\mathbf{v}$  at points  $\mathbf{r}$  with respect to an arbitrarily first chosen observer using time  $t$  will be a fundamental one in the above sense. The Lorentz transformation is then used to set up a distribution function in velocity space which shall be invariant in form as a function of velocity ; then the velocity is assumed as a function of the radius as given by (1) from the Newtonian transformation, and the resulting space-distribution function is presented as a fundamental system of equivalent observers.

1.5. In order to discover whether such a system of observers can continue to remain equivalent without constraints on their motion, the transformation (Lorentz) of acceleration is set up and an attempt made to find an expression for the acceleration which shall be invariant in form as a velocity function in the fundamental system of observers. Because this expression turns out to be zero everywhere if it is zero at time  $t=0$ , it is asserted that therefore the fundamental system will remain unaccelerated even without restraints and independently of any gravitational attractions between the observers.

1.6. Finally a test-particle is introduced. It is not clearly stated whether an observer is allowed to attach himself to the test particle, because it is assumed sometimes that this test particle is not a member of the fundamental system of equivalent observers. No details are given how the distance of the test particle is to be measured in the absence of an observer on the test particle to handle the reciprocal messages required for such a measurement. Milne would, however, probably regard this as a verbal quibble. The test particle is projected into the universe with a velocity different from that of its parent, and its motion is found to obey a cosmic law of gravitation, automatically, that will on the one hand almost coincide with Newtonian gravitation and on the other hand account for cosmic radiation.

1.7. Later, Milne<sup>(2)</sup> seems to have realised that a discrete set of observers, or particles, can never form a fundamental system of equivalent observers, and instead substituted the terminology "smoothed out substratum," with the idea that where a discrete system failed, a continuum without granulation, and having infinities only at infinity (where they can conveniently be ignored) may succeed. The pragmatic point of view, upon

which Milne seems at first to have prided himself, was also discreetly thrown overboard, for how can any logical definition of time measures be set up by comparing the clocks belonging to observers who have somehow got themselves smoothed out and lost their particularity in the substratum ?

*Criticism of the Axiomatic Basis of Milne's Theory.*

2.1. Milne's claim to have based world geometry on actually conceivable measurements has been adequately refuted by Dingle<sup>(3)</sup>. It will be as well to restate the argument here for the sake of emphasis. Milne's universe includes an infinity of particles moving with speeds approaching that of light at the confines of the universe. Synchronization of clocks as between ourselves and equivalent observers imagined as present in the distant nebulae involves in principle the exchange of wireless signals in both directions within a sufficiently short time to be sure that the observers' clocks have remained identical. According to Milne<sup>(4)</sup> we accept the two clocks as identical if, for example, the clock at B reads B.C. 5000 at A.D. 2000 on A's time, and again reads A.D. 9000 at A.D. 16,000 on A's time, even if the stellar observer was impish enough to change his time-piece while the terrestrial observer was otherwise engaged. It is surely impossible to set up a world geometry on the basis of time measures unless the possibility is admitted of some material permanency of exactly the type needed for rigid length-measuring rods. A clock, whether mechanical or atomic, which can safely be assumed to remain self-consistent throughout millions of years is no more conceivable, and no less conceivable, than a meter scale that will remain self-consistent under transport over millions of kilometers. Milne objects to the length measures, but swallows the time measures without hesitation.

2.2. Milne's later work requires, in addition to temporal synchronization, also the assignment of a definite epoch to the moment of creation : an absolute zero of time which apparently refers to a definite event described as the explosion of the universe out of a mathematical singularity in the density of the distribution function. In fact, every fundamental observer is expected to be able to locate this zero on his clock, and only then will all the clocks be fully synchronized and in step. This absolute zero of time is indeed included in Milne's list of immediately observable quantities, via the state of evolution of the nebulae as a function of the Doppler shift. Absolute time is supposed to be deducible from the state of evolution of the nebulae ! It is a remarkable form of measurement, and will naturally give results that depend upon the particular theory of evolution that happens to be favoured. This will not do for physicists, who have been measuring length and time even before the existence of the nebulae was known.

2.3. Milne claims to have reduced the postulate of the constancy of the velocity of light to a convention ; a matter of definition of length in terms of time. This attitude towards the postulate has long been



generally accepted, and has nothing to do with Milne's theory ; indeed, it is inconsistent with Milne's assertion that the flatness or otherwise of the continuum is also a matter of convention. To see this most easily, imagine three mutually stationary observers A, B and C at distances large enough to make sure that Milne's indirect method of comparing clocks, or some similar method, will have to be used to co-ordinate their measurements. Let A perform the Milne operations required to measure the distances AB and AC, and also, by means of his theodolite, measure the angle CAB, and finally calculate the distance BC. There is absolutely no axiomatic guarantee, even on Milne's theory, that this measurement of BC will agree with B's measure of the same distance by the equivalent methods. Any discrepancy can only be taken care of by A recognizing that the appropriate geometry is not flat. If, on the other hand, he takes the flatness as axiomatic, he must be prepared to admit variations in the velocity of light ; they cannot both be accepted as axiomatic ; one or other must be admitted as an observational fact. The point of view of the physicists who have to make the measurements is that the velocity of light, and the wave-length of light, form more suitable standards than any assumption about the nature of space-time extrapolated to the outermost bounds of the visible universe.

Milne certainly seems to have had some misgivings about this. He has<sup>(5)</sup> discovered that if the geometry is assumed to be flat even for an accelerated observer, it is necessary to assume that the cross-country light track, such as between B and C in the above example, must be curved relative to A when A is accelerated relative to B or C. It is difficult, however, to understand how a line BC could be curved with respect to A and yet straight to both B and C, if all the observers in question are using flat geometry. Either A must use a curved geometry, or else he must admit himself unable to assign unique co-ordinates to B or C. Milne has not presented any explicit transformations in flat space that would achieve this, but optimistically asserted that "it may be left for future research."

### *Criticism of Milne's Transformation Theory.*

3.1. In § 1.4 attention was drawn to the fact that Milne used invariance with respect to the Newtonian transformation of velocities to find the velocity as a function of position and time in his fundamental equivalence, and then inserted this result into the Lorentz-invariant velocity distribution function. A consistent procedure would at least have used a velocity function invariant for Lorentz transformations rather than Newtonian. Locally they make no difference, but Milne's applications are far from local.

Writing for the velocity in one dimension  $f(r, t)$  we have for the Lorentz transformation

$$f(r-r_0, t) = \frac{f(r, t) - f(r_0, t)}{1 - f(r, t)f(r_0, t)/c^2} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (2)$$



Seek a solution of the form  $f(r, t) = F(t)\phi(r)$ , . . . . . (3)

then 
$$\phi(r - r_0, t) = \frac{\phi(r) - \phi(r_0)}{1 - F(t)^2 \phi(r)\phi(r_0)/c^2}, \quad . . . . . (4)$$

which requires  $F(t)\phi(r) = c \tanh \alpha r \quad . . . . . (5)$

and so  $v = c \tanh \alpha r. \quad . . . . . (6)$

In a later series of papers Milne<sup>(6)</sup> noticed that his work reduced to pure Newtonian dynamics on the substitution of  $\tau$  for  $t$  where

$$\tau = t_0 \ln(t/t_0) + t_0, \quad . . . . . (7)$$

provided that all other measures deduced from time measures undergo the corresponding transformation. He then deduced a velocity function in terms of the dynamic time  $\tau$  that was practically identical with (6). This would seem to indicate that his kinematic time  $t$  is simply a mistake.

3.2. In any case the three-dimensional analysis is quite inadequate for a rigorous discussion of these problems, and so long as Milne refuses to use four-dimensional analysis the issues will remain nebulous and confused in the extreme.

The correct mathematical method for discovering Lorentz invariants is of course by means of four-dimensional vector analysis.

Introduce the unit vectors<sup>(7)</sup>

$$\mathbf{i}_0, \mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3, \quad . . . . . (8)$$

where  $\mathbf{i}_0 \cdot \mathbf{i}_0 = 1, \mathbf{i}_s \cdot \mathbf{i}_u = -\delta_{su}, s, u = 1, 2, 3, \quad . . . . . (9)$

then write  $\mathbf{R} = ct\mathbf{i}_0 + x\mathbf{i}_1 + y\mathbf{i}_2 + z\mathbf{i}_3, \quad . . . . . (10)$

and 
$$\mathbf{V} = c \frac{dt}{ds} \mathbf{i}_0 + \frac{dx}{ds} \mathbf{i}_1 + \frac{dy}{ds} \mathbf{i}_2 + \frac{dz}{ds} \mathbf{i}_3, \quad . . . . . (11)$$

where  $ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2. \quad . . . . . (12)$

Then the invariance of  $\mathbf{V}$  as a function  $\mathbf{R}$  is assured if and only if  $\mathbf{V}$  is, in four dimensions, a linear vector function of  $\mathbf{R}$ , thus

$$\mathbf{V} = \mathbf{T} \cdot \mathbf{R}, \quad . . . . . (13)$$

where  $\mathbf{T}$  is a mixed tensor, or a dyadic, having "dimensions" reciprocal length.

The most general transformation that preserves isotropy in the three-space dimensions is expressible in the form

$$\mathbf{V} = \mathbf{A}\mathbf{R} - \mathbf{B}\mathbf{N}, \quad . . . . . (14)$$

where  $\mathbf{A}$ ,  $\mathbf{B}$  are scalar invariants and  $\mathbf{N}$  is a null vector with the same projection in three-space as the vector  $\mathbf{R}$ . If  $\mathbf{B}$  were taken zero, it would mean that we had assumed isotropy not merely in three-space, but in four dimensions, which is more than the problem warrants.

Expressed in component notation (14) is simply

$$\left[ c \frac{dt}{ds}, \frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds} \right] = \mathbf{A}[ct, x, y, z] - \mathbf{B}[r, x, y, z]. \quad . . . . . (15)$$

The first, time-like, component now reads,

$$\frac{c}{\sqrt{c^2 - v^2}} = A c t - B r \quad . \quad . \quad . \quad . \quad . \quad . \quad (16 a)$$

and the last three space-like components can be collected into the form, using  $\mathbf{v}$ ,  $\mathbf{r}$  as the ordinary vectors of velocity and radius,

$$\frac{\mathbf{v}}{\sqrt{c^2 - v^2}} = (A - B) \mathbf{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16 b)$$

When  $t=0$  we have  $\mathbf{v} = c(1 - A/B) \mathbf{r}/r, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$

so that we may write, for convenience,

$$A/B = 1 - v_0/c, \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and finally

$$\mathbf{v} = \frac{v_0 \mathbf{r}}{r + t(v_0 - c)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

This is a differential equation for  $r$  in terms of  $t$ , and integrates to give

$$c t = r [1 - (r_0/r)^{c/v_0}], \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

showing that the value of  $r$  at  $t=0$  is arbitrary,  $r_0$ .

Only with  $r_0=0$  does this result reduce to Milne's particular solution. Alternatively, if  $B$  is assumed zero from the outset we also recover (1) from (16a) and (16b), but if we put  $t=0$  in (16a) in this case, either  $A$  must be infinite or  $v=c$ . To retain Milne's result it is then necessary to assume that  $A$  also has an infinite singularity at the moment of creation.

Here, however, we abandon the special restrictions needed to get Milne's result, and notice that from the exact kinematical requirements the fundamental system of particles can start from any desired radius with any desired velocity. And this arbitrariness is not merely a shift of time origin, (a) because the motion projected backwards to times earlier than  $t=0$  will reach  $r=0$  only at  $t \rightarrow -\infty$  unless of course  $v=c$ , and (b) because while the motion will be uniform if  $r_0=0$ , it is not otherwise uniform at all. The dependence of  $\mathbf{v}$  on  $\mathbf{r}$  is in general

$$\mathbf{v} = \frac{v_0 \mathbf{r}/r}{1 - (1 - v_0/c)[1 - (r_0/r)^{c/v_0}]} \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

This is zero at  $r=0$ , has the value  $v_0$  at  $r_0$  and approaches the limit  $c$  at  $r \rightarrow \infty$ .

3.3. The situation with regard to acceleration on Milne's theory is still worse. He writes down the Lorentz transformation equations for the three components of ordinary acceleration induced by a relative velocity between two observers and finds a particular solution of the functional equations which suits his particular purposes. No proof is given, nor even a hint that one is necessary, that the solution is general.

The general procedure is as follows. Write for the four-vectors representing acceleration, velocity and position

$$\mathbf{R} \sim [ct, x, y, z], \mathbf{V} \sim \left[ c \frac{dt}{ds}, \frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds} \right]$$

$$\mathbf{F} \sim \left[ c \frac{d^2t}{ds^2}, \frac{d^2x}{ds^2}, \frac{d^2y}{ds^2}, \frac{d^2z}{ds^2} \right], \dots \dots \dots (22)$$

and let  $\mathbf{F} = \Phi \cdot \mathbf{V} + \Psi \cdot \mathbf{R}, \dots \dots \dots (23)$

where  $\Phi$  and  $\Psi$  are dyadics whose components are functions of invariants formed from  $\mathbf{R}$ , with appropriate dimensional characteristics. This is a vector equation whose form will automatically be invariant under all Lorentz transformations exactly as demanded by Milne. The most general forms for  $\Phi$  and  $\Psi$  satisfying the condition that no preferential direction exists other than  $\mathbf{V}$  itself, which in this connection is to be regarded as a provisionally independent variable, are easily seen to reduce to

$$\Phi = \phi_{11} \frac{\mathbf{R}\mathbf{R}}{R^3} + \phi_{12} \frac{\mathbf{R}\mathbf{V}}{R^2} + \phi_{21} \frac{\mathbf{V}\mathbf{R}}{R^2} + \phi_{22} \frac{\mathbf{V}\mathbf{V}}{R} \dots \dots (24)$$

$$\Psi = \psi_{11} \frac{\mathbf{V}\mathbf{V}}{R^2} + \psi_{12} \frac{\mathbf{V}\mathbf{R}}{R^3} + \psi_{21} \frac{\mathbf{R}\mathbf{V}}{R^3} + \psi_{22} \frac{\mathbf{R}\mathbf{R}}{R^4}, \dots \dots (25)$$

where  $\phi_{\alpha\beta}$  and  $\psi_{\alpha\beta}$  are functions of  $c\mathbf{R} \cdot \mathbf{V}/R$  alone.

The denominators are so chosen as to reduce all terms in  $\mathbf{F}$  to the same dimensional character, and the  $\phi$ 's and  $\psi$ 's are all dimensionless invariants. Of course

$$R^2 = c^2t^2 - x^2 - y^2 - z^2 \dots \dots \dots (26)$$

and  $V^2 = (c^2 - v^2)(dt/ds)^2 = 1. \dots \dots \dots (27)$

It is easy to show next that

$$\mathbf{F} = \frac{\mathbf{R}}{R^2} \psi(\mathbf{R} \cdot \mathbf{V}/R) + \frac{\mathbf{V}}{R} \phi(\mathbf{R} \cdot \mathbf{V}/R), \dots \dots \dots (28)$$

where  $\psi$  and  $\phi$  are functions defined in terms of  $\phi_{\alpha\beta}$  and  $\psi_{\alpha\beta}$ .

Clearly this result is still restricted by symmetry as between space and time. To relax it we can add any null vector whose space projection coincides with the space projection of either  $\mathbf{V}$  or  $\mathbf{R}$ , so that the most general transformation that preserves isotropy in three-space is

$$\mathbf{F} = \frac{-\lambda \mathbf{N}_1}{R^2} \psi + \frac{\mathbf{V} - \gamma \mathbf{N}_2}{R} \phi, \dots \dots \dots (29)$$

where  $\mathbf{N}_1 = r\mathbf{i}_0 + x\mathbf{i}_1 + y\mathbf{i}_2 + z\mathbf{i}_3 \sim [r, x, y, z] \dots \dots \dots (30)$

and  $\mathbf{N}_2 = \frac{dr}{ds} \mathbf{i}_0 + \frac{dx}{ds} \mathbf{i}_1 + \frac{dy}{ds} \mathbf{i}_2 + \frac{dz}{ds} \mathbf{i}_3 \sim [\dot{r}, \dot{x}, \dot{y}, \dot{z}] \frac{dt}{ds} \dots \dots (31)$

are null vectors whose space-projections coincide with the space projections of  $\mathbf{R}$  and  $\mathbf{V}$  respectively, and  $\lambda$  and  $\gamma$  are arbitrary scalar invariants like  $\psi$  and  $\phi$ .

Writing (29) out in components we have

$$\left[ c \frac{d^2 t}{ds^2}, \frac{d^2 x}{ds^2}, \frac{d^2 y}{ds^2}, \frac{d^2 z}{ds^2} \right] = \frac{\psi}{R^2} \{ [ct, x, y, z] - \lambda [r, x, y, z] \} \\ + \frac{\phi}{R} \left\{ \left[ c \frac{dt}{ds}, \frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds} \right] - \gamma \left[ \frac{dr}{ds}, \frac{dx}{ds}, \frac{dy}{ds}, \frac{dz}{ds} \right] \right\}. \quad (32)$$

The first, time-like component then gives

$$c \frac{d}{ds} \left( \frac{dt}{ds} \right) = \frac{\psi}{R^2} (ct - \lambda r) + \frac{\phi}{R} (c - \gamma dr/dt) \frac{dt}{ds} \\ \text{or} \quad \frac{d}{dt} \left( \frac{dt}{ds} \right) = \frac{\psi}{R^2} (t - \lambda r/c) \frac{ds}{dt} + \frac{\phi}{R} (1 - \gamma dr/cdt). \quad (33)$$

The three last combine to give

$$\left( \frac{dt}{ds} \right)^2 \mathbf{g} = \frac{\psi}{R^2} (1 - \lambda) \mathbf{r} + \frac{\phi}{R} (1 - \gamma) \mathbf{v} \frac{dt}{ds} - \frac{dt}{ds} \mathbf{v} \frac{d}{dt} \left( \frac{dt}{ds} \right), \quad (34)$$

where  $\mathbf{g}$  has been written for the ordinary acceleration vector and we have used the relation

$$\frac{d^2 x_\alpha}{ds^2} = \left( \frac{dt}{ds} \right)^2 \frac{d^2 x_\alpha}{dt^2} + \left( \frac{dt}{ds} \right) \frac{dx_\alpha}{dt} \frac{d}{dt} \left( \frac{dt}{ds} \right). \quad (35)$$

Using (33) in the last term of (34) we at once obtain

$$\left( \frac{dt}{ds} \right)^2 \mathbf{g} = \frac{\psi}{R^2} [(1 - \lambda) \mathbf{r} - (1 - \lambda r/c) \mathbf{v}] \\ + \frac{\phi}{R} \left[ (1 - \gamma) - \left( 1 - \gamma \frac{dr}{cdt} \right) \right] \left( \frac{dt}{ds} \right) \quad (36)$$

$$= \frac{\psi}{R^2} (1 - \lambda) \mathbf{r} - \left\{ \frac{\psi}{R^2} (t - \lambda r/c) + \gamma \frac{\phi}{R} \left( 1 - \frac{1}{c} \frac{dr}{dt} \right) \frac{dt}{ds} \right\} \mathbf{v}, \quad (37)$$

or finally

$$\mathbf{g} = \mathbf{r} \frac{\psi}{R^2} (1 - \lambda) (c^2 - v^2) \\ - \mathbf{v} \left\{ \frac{\psi}{R^2} (t - \lambda r/c) (c^2 - v^2) + \gamma \frac{\phi}{R} \left( 1 - \frac{1}{c} \frac{dr}{dt} \right) (c^2 - v^2)^{\frac{1}{2}} \right\}. \quad (38)$$

Only if in particular  $\lambda$  and  $\gamma = 0$ , so that  $\mathbf{F}$  is symmetrical as between space and time will this reduce to Milne's result.

$$\mathbf{g} = (c^2 - v^2) (\mathbf{r} - \mathbf{v}t) \psi / R^2. \quad (39)$$

It is, in fact, certainly unnecessary to suppose that  $\lambda$  and  $\gamma$  are zero from any strict requirement regarding the invariance of  $\mathbf{F}$  or the isotropy of three-space; in general they may be any scalar functions whatever of  $\mathbf{R}$ .

The significance of these results is that Milne's claim to have found a unique set of trajectories based on his particular  $\mathbf{g}$  formula is false. With the freedom of choice represented by the general solution (38) almost any trajectory becomes kinematically possible, and a much stronger principle of selection will be required to specify any particular set as representing actual motion: in other words some law of gravitation independent of mere kinematical considerations.



*Review of Published Comments on Milne's Theory.*

4.1. T. Lewis<sup>(8)</sup> deduced Milne's acceleration transformation formula from Minkowski's four-dimensional analysis but failed to relax the isotropy between space and time implied by that analysis.

4.2. C. Gilbert<sup>(9)</sup> also derived Milne's trajectories from special relativity by means of a co-ordinate transformation equivalent to a regraduation of Milne's clocks.

4.3. McCrea<sup>(10)</sup> also obtained the acceleration formula by means of group theory in three dimensions.

4.4. McVittie<sup>(11)</sup> gave all Milne's observational results in terms of general relativity theory, and later<sup>(12)</sup> re-affirmed that nebular shift and nebular count data demanded a hyperbolic universe, thus destroying Milne's claim that his theory had obviated the necessity for any universal constant other than the velocity of light.

4.5. Hubble and Tolman<sup>(13)</sup>, Silberstein<sup>(14)</sup>, Shapley<sup>(15)</sup> and McVittie<sup>(16)</sup> have discussed the observational data from the point of view of the expanding universe, and the general impression the reader of these papers gathers is that the interpretation of data is still a matter of debate, but that whatever model of the universe is eventually agreed to fit the facts, it will be capable of being handled by the general theory of relativity in a satisfactory manner.

4.6. In fact Robertson<sup>(17)</sup> has elaborated Milne's model itself from the general relativity point of view in an entirely satisfactory manner.

4.7. Dingle's criticisms of Milne's axioms have already been mentioned in § 2.1 above. There has also been some amusing correspondence in the columns of 'Nature' recently from which one would gather that Milne is beginning to withdraw from his original philosophical position by belittling the importance of criticisms of the literary phraseology with which he has "from time to time" chosen to dress up his mathematics. In the present paper it is not only his phraseology but his mathematics that are shown to be inadequate.

4.8. There has also been an exchange of papers recently on the acceleration transformation of Milne's. Wilson<sup>(18)</sup> objected to the formula which appears as (39) in this paper, because it is not homogeneously four-dimensionally invariant in form. Newing<sup>(19)</sup> supported Milne's formula by showing that although not four-vector in form, it is nevertheless homogeneous and Lorentz invariant. Whitrow<sup>(20)</sup> also showed that it could be put into the form of the space-part of a four-vector. The argument given in the present paper of course agrees with Whitrow and Newing; the equation is deducible from a general four-vector equation as a special case. The result finally is, however, to undermine Milne's interpretation completely by proving that his formula is not the general solution of the problem.

*Milne's Criticism of "Current Relativity."*

5.1. Milne <sup>(21)</sup> objects to the circumstances that general relativity makes space depend upon matter distribution in a manner that cannot be mapped until space is first assumed. No observational methods have been given by means of which the metric of space can be chosen without a previous knowledge of the distribution of matter. Milne states that this involves a circular argument which he claims to have avoided ; his axioms leading to a space metric independent of what matter may be present.

In § 2.3 above it was, however, shown that Milne's clock-cum-theodolite method of setting up a world geometry does not automatically lead to the flat space assumed by Milne, but may quite well lead the observer to discover curvature in the universe. Here it will be shown that not only does Milne's theory fail to define a flat geometry, but it fails to set up an affine co-ordinate system of any kind.

Consider the three mutually stationary observers A, B and C as before, but now suppose a stellar mass D to pass slowly across the space between A and the other two. The gravitational field of D will perturb the light signals between A, B and A, C in an entirely determinate manner, from the general relativity point of view. But, from the Milne point of view, the signals are a matter of definition and cannot be perturbed. The theodolite will register slight lateral movements of B and of C in turn as D passes across A's field of view, and by definition these lateral movements are real. If they are not real, Milne must provide additional criteria beyond the information derivable from A's clock and theodolite ; he must, in fact, assume the nature of the space and abandon the light signal as a final definition, regarding the gravitational field as perturbing the signals in the usual Newtonian manner. If, however, we accept Milne's own statement of his theory, the nature of the metric is determined directly from the clock-theodolite measurements alone, by means of light signals. This would involve him in a dilemma if he still assumes the result to be flat.

The situation becomes even more obvious if we allow the mass D to be in such a position that its gravitational field exerts a lens-like action on the light signals from B or C, producing an image of the particle B at a point between A and D which, to all optical appearances as revealed by Milne's measurements, is just as real as the particle B itself, and is by definition actually at the measured position of B relative to A.

The fundamental weakness of Milne's theory is, as was pointed out earlier, that no theory of measurement can safely be built up by means of comparisons made over vast distances in space or hoary eons of time, when there is no guarantee of absolute uniformity over these great distances and immense times. It is exactly the absence of this uniformity in fact that prevents any such wild extrapolations as Milne proposes from being of any practical use.

5.2. Milne has drawn a distinction between transformations of co-ordinates and transformations of observers <sup>(23)</sup>, and claims that his theory is superior to current relativity in that it is only necessary to arrange for

the functional form of variables defining the motion of any particle to remain invariant as between individual equivalent observers. He also states that general relativity regards a law of nature as a condition of conservation of values of a tensor, in particular the value zero, rather than a conservation of functional form. In one place <sup>(24)</sup> he states that restricted relativity starts by establishing the conservation of form of the quadratic expression for the square interval, and then general relativity loses sight of the form and seeks only to conserve the value of  $ds^2$ . But this can hardly be maintained. As the writer has pointed out in another connection <sup>(25)</sup>, the Lorentz transformation at any point in Riemannian space does not change the metric coefficients locally, and does, even in Milne's sense, conserve both form and value.

Milne seems to have been confusing two distinct concepts, namely co-ordinate transformation and space-mapping respectively, and it will be as well to discuss these briefly here. The co-ordinate transformation can be represented as follows:—Let a given vector field  $\mathbf{A}$  be represented on two alternative co-ordinate reference systems  $\mathbf{i}_\alpha$  and  $\mathbf{i}_\alpha'$ ; then

$$A^\alpha \mathbf{i}_\alpha = A^{\alpha'} \mathbf{i}_\alpha', \quad \dots \dots \dots (40)$$

because it is the same vector on either system, and where

$$\mathbf{i}_\alpha' = \phi^\beta_\alpha \mathbf{i}_\beta, \quad \mathbf{i}_\alpha = \phi^\beta_\alpha \mathbf{i}_\beta' \quad \dots \dots \dots (41)$$

represents the co-ordinate transformation. Comparing (41) and (40) we see that

$$A^{\alpha'} = A^\beta \phi^\alpha_\beta, \quad A^\alpha = A^{\beta'} \phi^\alpha_{\beta'}, \quad \dots \dots \dots (42)$$

the components of the given vector, are different functions of the co-ordinates when the latter are changed. Milne now takes a different vector,  $\mathbf{B}$  which has the same components in the new co-ordinates as  $\mathbf{A}$  has in the original co-ordinates:

$$B^{\alpha'} = A^\alpha \quad \dots \dots \dots (43)$$

If now  $B^\alpha$  are the components of  $\mathbf{B}$  corresponding to the original co-ordinates, we shall have, using (43) and (42),

$$B^\alpha \mathbf{i} = B^{\alpha'} \mathbf{i}_\alpha' = A^\alpha \mathbf{i}_\alpha' = A^\alpha \mathbf{i}_\beta \phi^\beta_\alpha, \quad \dots \dots \dots (44)$$

so that

$$B^\alpha = \phi^\alpha_\beta A^\beta. \quad \dots \dots \dots (45)$$

This is Milne's functional transformation formula. It represents the mapping procedure by which the vector field  $\mathbf{B}$  is obtained from the vector field  $\mathbf{A}$ . Every co-ordinate transformation induces a congruent mapping, and *vice versa*, thus proving that Milne's functional transformations are in every mathematical respect equivalent to the ordinary transformations of restricted relativity.

### *The Axioms of General Relativity.*

6.1. There remains only one objection of Milne's that does not appear to have been answered elsewhere, namely that general relativity presupposes a system of co-ordinates arbitrarily set up throughout the universe by



means of which every event can in principle be located, but gives no definite indication how any given observer is expected to set up such a system. The real significance of this would appear to have been lost to the Milne school; namely, that general relativity presents a programme by means of which any arbitrary assignment of co-ordinates can be employed, without making any observable difference to actual phenomena or the laws of nature, provided only that the assignment of co-ordinates is properly affine. There is, in general relativity, no need to set up a prescribed procedure for experimental assignment of co-ordinates before we can arrive at an understanding of dynamical phenomena. After all, this is a tremendous achievement, for if we had to await a complete theory of measurement before we could set up a theory of dynamics, our position would be quite hopeless.

6.2. The required self-consistency, or affine nature, of co-ordinate assignments does, however, involve important restrictions upon our experimental procedures, and at once rules out the methods imagined by Milne, where distances are measured in terms of light signals traversing inter-stellar space without correction for perturbations. As we have pointed out, his method will lead to a non-affine co-ordinate system in practice, unless there is no gravitational field anywhere. Self-consistency demands at least that no comparisons be taken for granted as between either clocks or scales separated either by great distances or long times. Any attempt by an observer to make measurements outside of his own immediate surroundings must be regarded as extrapolation, and extrapolation is dangerous.

Instead we must appeal to the principle of connectivity, which states that it is possible for any observer  $S$  to compare, with his own, the unit reference vectors used almost simultaneously by any other observer  $S'$  in a position not too far distant from  $S$  at the time of comparison.

We can now populate the universe, in imagination, with free test-particle observers provided with conventional measuring devices, one observer in each four-dimensional element of space-time, and at once set up the fundamental equations of differential geometry by means of their mutual comparisons of reference units. This procedure is equivalent with a well-known mathematical theorem that any Riemannian space can be built up by means of a family of local flat tangent spaces. I have described the procedure in detail elsewhere <sup>(26)</sup>, and it is sufficient here to point out that it places general relativity on a secure pragmatic basis: it ensures that world geometry and the co-ordinate system become results of well-defined operational procedures of comparison between neighbouring members of any system of observers populating the universe. To form a satisfactory basis of general relativity these observers do not have to be "equivalent" in the Milne sense. The great achievement of the Einstein theory is, in fact, that there is no privileged set of select "equivalent" observers. No matter who looks at the universe, he will see essentially the same external world. This point of view is vastly



more satisfying than Milne's, and the methods of analysis provided by the Riemannian geometry are incomparably more powerful than those used by Milne to handle his particular system of observers.

In conclusion, there seems no possible doubt that Milne's kinematical relativity, and all related efforts to establish a theory of the universe which refuses to make use of general relativity, are justifiable only as stimulants of interesting discussion, and sources of amusement to their authors.

### References.

- (1) Milne, 'Relativity, Gravitation and World Structure,' (Oxford, 1935).
- (2) Milne, Proc. Roy. Soc., A, 156 (1936) ; A, 160 (1937).
- (3) Dingle, 'Philosophy,' xi. (Jan. 1936).
- (4) Milne, 'Relativity, Gravitation and World Structure,' p. 31, fig. 2 (Oxford, 1935).
- (5) Milne, 'Relativity, Gravitation and World Structure,' § 39 (Oxford, 1935).
- (6) Milne, Proc. Roy. Soc., A, 158, 159 (1937).
- (7) W. Band, Am. J. Phys. viii. (June, 1940).
- (8) T. Lewis, Phil. Mag. xx, p. 1092-1104 (Dec. 1935).
- (9) C. Gilbert, Phil. Mag. xxvii. (May, 1939).
- (10) W. H. McCrea, Proc. Roy. Irish Acad. xlv. A, pp. 23-30 (1938).
- (11) G. C. McVittie, Roy. Ast. Soc. M. N. xciv. pp. 476-483 (March, 1934).
- (12) G. C. McVittie, Roy. Ast. Soc. M. N. xcvii. pp. 163-167 (Jan. 1937) ; *Z. Astroph.* 13, v. pp. 312-320 ; 14, iv. pp. 274-284 (1937).
- (13) Hubble and Tolman, *Astroph. J.* lxxxii. pp. 302-337 (1935).
- (14) Silberstein, Roy. Soc. Canada Trans. xxix. (3), pp. 1-4 (May, 1935).
- (15) Shapley, Nat. Acad. Sci. Proc. xxiv. pp. 148-154, 282-287 (1938).
- (16) G. C. McVittie, 'Observatory,' p. 61 (Aug. 1938).
- (17) H. P. Robertson, *Astroph. J.* lxxxii. pp. 284-301 (1935), and lxxxiii. pp. 187-201 (1936).
- (18) Wilson, Phil. Mag. xxxv. p. 241 (1944).
- (19) Newing, Phil. Mag. xxxvi. p. 113 (1945).
- (20) Whitrow, Phil. Mag. xxxvi. p. 170 (1945).
- (21) Milne, 'Relativity, Gravitation and World Structure,' §§ 506, 514 (Oxford, 1935).
- (22) Einstein, 'Science,' lxxxiv. pp. 506-507 (Dec. 1936).
- (23) Milne, 'Relativity, Gravitation and World Structure,' §§ 377-8 (Oxford, 1935).
- (24) Milne, 'Relativity, Gravitation and World Structure,' footnote to p. 94 (Oxford, 1935).
- (25) W. Band, Phys. Rev. lxi. p. 705 (1942).
- (26) W. Band, Phys. Rev. lxi. p. 702 (1942).

### LVI. On the Theory of Boundary Perturbations.

By G. D. WASSERMANN,

The Mining and Technical College, Wigan\*.

[Received July 3, 1945.]

It is well known that only a few of the partial differential equations of mathematical physics allow an exact solution by the method of separation

\* Communicated by T. G. Cowling.

of variables. Quite often, however, one requires the solution of a wave equation which differs but slightly from another equation whose exact solution is known. Schroedinger established an elegant perturbation calculus which does justice to many problems of this type. Schroedinger's calculus deals predominantly with infinite boundaries. The boundary conditions are usually such that any surface integrals that would appear through the application of Green's theorem simply disappear, which causes all operators to be Hermitian. In a perturbation theory dealing with systems enclosed by finite surfaces surface integrals will frequently emerge, with a consequent loss of the Hermitian property, which can only be artificially retained by the introduction of "improper" surface operators (Fuchs, 1940). Moreover, two distinct types of perturbation are possible, one arising from an internal change of potential of the system (*e. g.* the addition of spin terms) the other from a slight change of boundaries or boundary conditions or both. This paper aims at establishing a consistent finite boundary perturbation theory.

Of the previous work in this field we mention the papers by Froehlich (1938), Cabrerra (1938), Brillouin (1937), Feshbach and Clogstone (1941), Feshbach (1944), and Fuchs (1940). Feshbach and Clogstone's attempt comprises and generalises the work of the earlier writers. The authors, however, introduce the boundary perturbations by means of perturbation potentials in a way that leads to an indeterminacy of the higher order perturbation terms.

Our investigation leads to results which are substantially in agreement with those of Feshbach, who derived them by means of an integral equation involving a Green's function. Our approach does not depend on the explicit use of a Green's function nor on the introduction of improper surface operators. Our work depends on a species of differential equations. They can be solved by a series solution, which leads to Feshbach's results. These series solutions converge, however, very slowly for the higher eigenvalues and eigenfunctions. The principal advantage of our work arises from the fact that one can also obtain solutions of these differential equations which do not depend on the use of infinite series. We are thinking of Temple's method (Temple, 1938) which, by an appropriate breaking off of the series used there, may give a useful rough approximation. The use of a differential analyser is indicated in simple two-dimensional problems. The application of these methods to the solution of our differential equations requires still further research.

Finally, it should be noticed that although we restrict our theory to the Schroedinger equation, whatever is said holds with but slight modifications for many other types of linear differential equation to which our theory can be extended.

### (1) *The Perturbation of Boundary Conditions.*

Let the eigenvalues  $K_{n0}$  and eigenfunctions  $\psi_{n0}$  of the wave equation

$$H^0 \psi_{n0} + K_{n0} \psi_{n0} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

be known inside a region  $D_0$  and on the surface  $S_0$  enclosing this region. It is assumed that  $H^0$  is a Hamiltonian operator of the form

$$H^0 = \nabla^2 + f(x),$$

where  $f(x)$  is a function of the co-ordinates, and that the  $\psi_{n0}$  obey homogeneous boundary conditions (B.C.) on  $S_0$  and are normalized to unity in  $D_0$ .

We confine our discussion to the following four types of B.C. for any wave function  $\psi$  on  $S$ :—

$$(I) \quad \psi = 0, \quad (II) \quad \frac{\partial \psi}{\partial n} = 0,$$

$$(III) \quad \frac{\partial \psi}{\partial n} = F(x)\psi, \quad (IV) \quad F(x) \frac{\partial \psi}{\partial n} = \psi,$$

where  $|F(x)|$  is small compared to unity.

Suppose now we wish to find a solution  $\psi_n$  of (1) satisfying either the same type of B.C. on a slightly perturbed surface  $S$  as  $\psi_{n0}$  does on  $S_0$ ; or we may demand that  $\psi_n$  should satisfy B.C. (III) or (IV) on the original surface  $S_0$  or the perturbed surface  $S$  when  $\psi_{n0}$  satisfies B.C. (II) or (I) respectively on  $S_0$ . This solution  $\psi_n$  and its eigenvalue  $K_n$  may be supposed derived from  $\psi_{n0}$  and  $K_{n0}$  by a perturbation process. We do not consider inhomogeneous B.C., but they can be dealt with in a similar manner.

Following Brillouin we introduce a displacement operator  $R$  transforming the point  $(x_i)$  on the unperturbed surface  $S_0$  into the point  $x_i + X(x_i)$  on the perturbed surface  $S$ .  $X(x)$  is a function of the co-ordinates. Hence

$$Rf(x) = f(x+X) = (1 + \mu^I + \mu^{II} + \dots + \mu^r + \dots)f(x), \quad \dots \quad (2)$$

where  $\mu^r$  is the component of  $r$ th order smallness of  $R$ , so that

$$\mu^I = \sum_{i=1}^3 X_i \frac{\partial}{\partial x_i} \quad \text{and} \quad \mu^r = \sum_{l+m \leq r} \frac{1}{l! m! (r-l-m)!} X_1^l X_2^m X_3^{r-l-m} \\ \times \frac{\partial^r}{\partial x_1^l \partial x_2^m \partial x_3^{r-l-m}} \quad \text{for } r > 1. \quad (3)$$

It is important to notice that  $R$  transforms any function on  $S_0$  into a function of the same form on  $S$ . Hence if the equations of  $S_0$  and  $S$  are

$$S_0(x) = 0, \quad S(x) = 0,$$

then, for  $x$  on  $S_0$ ,

$$RS(x) = S(x+X) = 0,$$

so that  $S(x)$  can be defined to be such that

$$RS(x) = S_0(x), \quad S(x) = R^{-1}S_0(x). \quad \dots \quad (4)$$

Next we expand  $\psi_n$  and  $K_n$  in the forms

$$\psi_n = \psi_{n0} + \psi_{n1} + \dots + \psi_{nr} + \dots \quad \dots \quad (5)$$

and

$$K_n = K_{n0} + K_{n1} + \dots + K_{nr} + \dots, \quad \dots \quad (6)$$

where  $\psi_{nr}$  and  $K_{nr}$  are the components of  $r$ th order smallness of  $\psi_n$  and  $K_n$  respectively.

Case (1).—Suppose  $\psi_{n0}$  satisfies B.C. (II) on  $S_0$ , while  $\psi_n$  satisfies B.C. (III) on  $S$ . Then on  $S$

$$F(x)\psi_n = -\frac{\partial\psi_n}{\partial n} = \frac{\nabla S \cdot \nabla\psi_n}{|\nabla S|}.$$

Since  $S(x) = R^{-1}S_0(x)$ , the last expression can be expanded in the form

$$\left\{ \frac{\partial}{\partial n_0} + \frac{\partial}{\partial n_1} + \frac{\partial}{\partial n_2} + \dots \frac{\partial}{\partial n_r} + \dots \right\} \psi_n,$$

where

$$\frac{\partial\psi_n}{\partial n_0} = \frac{\nabla S_0 \cdot \nabla\psi_n}{|\nabla S_0|},$$

and  $\partial/\partial n_r$  denotes the part of  $\partial/\partial n$  of  $r$ th order smallness. Hence on  $S_0$

$$R \left\{ \frac{\partial\psi_n}{\partial n_0} + \frac{\partial\psi_n}{\partial n_1} + \frac{\partial\psi_n}{\partial n_2} + \dots \right\} = R \{ F(x)\psi_n \}. \quad (7)$$

Substituting from (2) and (5) into (7), and separating terms of successive orders of smallness. ( $F$  is assumed to be of the first order.)

$$\frac{\partial\psi_{n1}}{\partial n_0} = F(x)\psi_{n0} - \mu^I \frac{\partial\psi_{n0}}{\partial n_0} - \frac{\partial\psi_{n0}}{\partial n_1}, \quad (8.1)$$

$$\frac{\partial\psi_{n2}}{\partial n_0} = F(x)\psi_{n1} - \mu^{II} \frac{\partial\psi_{n0}}{\partial n_0} - \mu^I \left( \frac{\partial\psi_{n1}}{\partial n_0} + \frac{\partial\psi_{n0}}{\partial n_1} \right) - \frac{\partial\psi_{n1}}{\partial n_1} - \frac{\partial\psi_{n0}}{\partial n_2}. \quad (8.2)$$

and so on. Since (8.2) involves  $\psi_{n1}$ , the B.C. for  $\psi_{n2}$  cannot be given explicitly until  $\psi_{n1}$  has been determined.

Case (2).—Let  $\psi_{n0}$  satisfy B.C. (I) on  $S_0$ , while  $\psi_n$  satisfies B.C. (IV) on  $S$ . Proceeding as before we obtain

$$R \left( F \left\{ \frac{\partial\psi_n}{\partial n_0} + \frac{\partial\psi_n}{\partial n_1} + \dots \right\} \right) = R\psi_n \quad (9)$$

for  $x$  on  $S_0$ . Separation into orders of smallness can immediately be effected as for (7). Thus for the first order

$$\psi_{n1} = F \frac{\partial\psi_{n0}}{\partial n_0} - \mu^I \psi_{n0} \quad (10)$$

on  $S_0$ ; and the equation of the  $r$ th order gives the value of  $\psi_{nr}$  on  $S_0$ . This concludes our discussion of the boundary values of the  $\psi_{nr}$  in some typical cases.

## (2) The Extended Schroedinger System for Nondegenerate States.

Suppose  $\psi_n$  satisfies

$$H\psi_n + K_n\psi_n = 0 \quad (11)$$

inside and on  $S$ , subject to one of the B.C. (I)–(IV). We assume an expansion of  $H$  in orders of smallness

$$H = H^0 + H^1 + H^2 + \dots + H^r + \dots \quad (12)$$

$H^r$  is supposed to represent a perturbation potential



Substituting (6), (5) and (12) into (11) and separating the parts of successive orders of smallness, we obtain equation (1) and the system

$$H^0 \psi_{n1} + K_{n0} \psi_{n1} = -(K_{n1} + H^1) \psi_{n0}, \quad \dots \quad (13.1)$$

$$H^0 \psi_{nr} + K_{n0} \psi_{nr} = - \sum_{\lambda=0}^{r-1} (K_{n, r-\lambda} + H^{r-\lambda}) \psi_{n\lambda}, \quad \dots \quad (13.r)$$

( $r=2, 3, \dots$ ).

Multiplying (13.r) by  $\bar{\psi}_{s0}^*$  and integrating over  $D_0$  we obtain, by using Green's theorem,

$$\begin{aligned} - \sum_{\lambda=0}^{r-1} \int_{D_0} \bar{\psi}_{s0} (K_{n, r-\lambda} + H^{r-\lambda}) \psi_{n\lambda} d\tau &= \int_{D_0} (\bar{\psi}_{s0} H^0 \psi_{nr} + K_{n0} \bar{\psi}_{s0} \psi_{nr}) d\tau \\ &= (K_{n0} - K_{s0}) \int_{D_0} \bar{\psi}_{s0} \psi_{nr} d\tau + \int_{S_0} \left( \bar{\psi}_{s0} \frac{\partial \psi_{nr}}{\partial n_0} - \psi_{nr} \frac{\partial \bar{\psi}_{s0}}{\partial n_0} \right) dS, \quad \dots \quad (14) \end{aligned}$$

( $s=1, 2, \dots$ )  
( $r=1, 2, \dots$ )

Define the matrix  $Y^r$  by

$$(s|Y^r|n) = \int_{S_0} \left( \psi_{nr} \frac{\partial \bar{\psi}_{s0}}{\partial n_0} - \bar{\psi}_{s0} \frac{\partial \psi_{nr}}{\partial n_0} \right) dS. \quad \dots \quad (15)$$

Then when  $n=s$  (14) reduces to

$$\sum_{\lambda=0}^{r-1} K_{n, r-\lambda} \alpha_\lambda = (n|Y^r|n) - \sum_{\lambda=0}^{r-1} \int_{D_0} \bar{\psi}_{n0} H^{r-\lambda} \psi_{n\lambda} d\tau, \quad \dots \quad (16)$$

( $r=1, 2, \dots$ )

where

$$\alpha_\lambda = \int_{D_0} \psi_{n\lambda} \bar{\psi}_{n0} d\tau.$$

Equations (13) and (16) in conjunction with the boundary conditions allow a complete formal solution of the problem.

Consider first B.C. case (2). Then  $\psi_{n0}=0$  on  $S_0$ , so that

$$(n|Y^r|n) = \int_{S_0} \psi_{nr} \frac{\partial \bar{\psi}_{n0}}{\partial n_0} dS. \quad \dots \quad (17)$$

Suppose that the functions  $\psi_{n1}, \psi_{n2}, \dots, \psi_{n, r-1}$  and the constants  $K_{n1}, K_{n2}, \dots, K_{n, r-1}$  have already been found. Then the value of  $\psi_{nr}$  on  $S_0$  can be found from the B.C. and so  $(n|Y|r)$  can be determined. Equation (16) can accordingly be used to calculate  $K_{nr}$ . This known, the right-hand side of (13.r) is a known function, so that  $\psi_{nr}$  is given as the solution of an inhomogeneous differential equation of the type

$$H^0 \phi + K_{n0} \phi = \mu(x). \quad \dots \quad (18)$$

The argument for B.C. case (1) is similar, save that (17) is replaced by

$$(n|Y^r|n) - \int_{S_0} \bar{\psi}_{n0} \frac{\partial \psi_{nr}}{\partial n_0} dS,$$

and  $\frac{\partial \psi_{nr}}{\partial n_0}$  is found from the B.C. For  $H^1=0$  and  $r=1$  (16) reduces to the equation for  $K_{n1}$  previously found by Froehlich.

---

\* The complex conjugate of  $\psi_{s0}$ .

One method of solution of (18) is based on the use of a Green's function. In fact Feshbach and Clogstone suggested this method when they succeeded in establishing (13) and (16) for  $r=1$  by a less systematic argument. Very few Green's functions are known in a closed form and a series expansion is preferable. A convergence proof for this expansion, though desirable, is usually extremely difficult; but in some cases convergence is sufficiently rapid to yield satisfactory numerical results. For the higher order eigenfunctions, which, though less important in Quantum Mechanics, play a prominent part in Acoustics, series expansion is inadequate, owing to slow convergence. This method of solution of (18) is fortunately not the only one available, and, as already mentioned, Temple's procedure and other approximate methods might prove more successful. As the application of these methods to (18) requires further careful research we confine ourselves in this paper to a discussion of the series solution.

We prefer to use a series expansion directly without intermediate resort to a Green's function. Inserting expansion

$$\psi_{nr} = \sum_s a_{sn}^r \psi_{s0} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

into (14) we obtain the transformation into matrix form, allowing an algebraic solution of the problem. For this purpose we define the matrices  $H^0$ ,  $H^t$ ,  $U^t$ , and  $K^t$  with components

$$\begin{aligned} (r|H^0|s) &= K_{r0} \delta_{rs}, \\ (r|H^t|s) &= \int_{D_0} \bar{\psi}_{r0} H^t \psi_{s0} d\tau, \\ (r|U^t|s) &= a_{rs}^t, \\ (r|K^t|s) &= K_{rt} \delta_{rs}. \end{aligned}$$

( $t=1, 2, \dots$ )

Equation (14) now transforms into the matrix equations

$$[U^r H^0] \equiv U^r H^0 - H^0 U^r = Y^r - \sum_{\lambda=0}^{r-1} U^\lambda K^{r-\lambda} - \sum_{\lambda=0}^{r-1} H^{r-\lambda} U^\lambda. \quad (20)$$

As the general solution is rather complicated we shall only deduce the practically important first- and second-order solutions. The  $(n, n)$  components give

$$K_{n1} = (n|Y'|n) - (n|H^1|n), \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$K_{n2} = (n|Y^2|n) - (n|H^2|n) - \sum_{p \neq n} (n|H'|p)(p|U'|n). \quad . \quad . \quad (22)$$

Equating  $(s, n)$  components ( $s \neq n$ ) we obtain

$$(s|U^1|n) = \frac{(s|Y'|n) - (s|H^1|n)}{K_{n0} - K_{s0}}, \quad . \quad . \quad . \quad . \quad . \quad (23)$$

$$\begin{aligned} (s|U^2|n) &= \frac{(s|Y^2|n) - (s|H^2|n)}{K_{n0} - K_{s0}} \\ &\quad - \sum_{p \neq n} \frac{(s|H'|p)((p|Y'|n) - (p|H'|n))}{(K_{n0} - K_{p0})(K_{n0} - K_{s0})} \\ &\quad - K_{n1} \frac{((s|Y'|n) - (s|H'|n))}{(K_{n0} - K_{s0})^2}. \quad . \quad . \quad . \quad . \quad (24) \end{aligned}$$

Hence 
$$K_{n2} = (n|Y^2|n) - (n|H^2|n) - \sum_{s \neq n} \frac{(n|H'|s)((n|Y'|s) - (s|H'|n))}{K_{n0} - K_{s0}} \quad (25)$$

$$\psi_{n1} = \sum_{s \neq n} \frac{(s|Y'|n) - (s|H'|n)}{K_{n0} - K_{s0}} \psi_{s0}, \quad (26)$$

and 
$$\begin{aligned} \psi_{n2} = & - \sum_{s \neq n} \frac{(s|H'|p)((p|Y'|n) - (p|H'|n))}{K_{n0}(-K_{p0})(K_{n0} - K_{s0})} \psi_{s0} \\ & - K_{n1} \sum_{s \neq n} \frac{(s|Y'|n) - (s|H'|n)}{(K_{n0} - K_{s0})^2} \psi_{s0} \\ & + \sum_{s \neq n} \frac{(s|Y^2|n) - (s|H^2|n)}{(K_{n0} - K_{s0})} \psi_{s0}. \quad (27) \end{aligned}$$

The corresponding equations in Feshbach's theory, on the other hand, are for case (1)

$$\psi_n = \psi_{n0} + \sum_{t \neq n} \frac{F_{tn}}{K - K_{t0}} + \sum_{s, t \neq n} \frac{F_{ts} F_{sn}}{(K - K_{t0})(K - K_{s0})} \psi_{t0} + \dots,$$

where  $K$  is a solution of

$$K = K_{n0} + F_{nn} + \sum_{t \neq n} \frac{(F_{tn})^2}{K - K_{t0}} + \dots$$

with

$$F_{tn} = - \int_S \psi_{t0} F \psi_{n0} dS,$$

and the last integral is taken over the perturbed surface  $S$ . Our own matrix elements are obtained by integration over the unperturbed surface  $S_0$ , and the perturbation appears in the integrand. In spite of this formal difference both theories yield the same result to the required degree of approximation. To obtain agreement with Feshbach's work we have of course to consider the simpler case  $H^1 = H^2 = H^n \dots = 0$ . Convergence difficulties for the high-order eigenfunctions will be equally marked in both theories.

It is useful to distinguish two cases.

(a) The B.C. of  $\psi_{n1}$  and  $\psi_{n0}$  on  $S_0$  are the same. Then the surface integrals vanish and the theory reduces to the ordinary perturbation formalism, save that all volume integrals must be taken over  $D_0$ .

(b) The B.C. of  $\psi_{n1}$  and  $\psi_{n0}$  on  $S_0$  differ. Equation (19) will now represent  $\psi_{n1}$  inside  $D_0$  but there are convergence difficulties on  $S_0$ . We notice that in case (1)  $\psi_{n1}$ , when obtained from (19), will have a non-uniformly convergent normal derivative on  $S_0$ , since  $\frac{\partial \psi_{n0}}{\partial n_0}$  and  $\frac{\partial \psi_{n1}}{\partial n_0}$  obey different

B.C. on  $S_0$ . To find  $\frac{\partial \psi_{n2}}{\partial n_0}$  from (8.2) we substitute  $\psi_{n1}$  in series form into the right-hand side. In the term  $\mu^1 \frac{\partial \psi_{n1}}{\partial n_0}$  we substitute  $\frac{\partial \psi_{n1}}{\partial n_0}$  from (8.1). The term  $\frac{\partial \psi_{n1}}{\partial n_0}$  can usually be split into expressions in which a linear operator operates on  $\psi_{n1}$  and  $\frac{\partial \psi_{n1}}{\partial n_0}$ , and we can use the same substitutions

as before. Each step must, however, be carefully examined for convergence. A similar procedure is as a rule inapplicable to case (2), since the series for  $\psi_{n1}$  is itself not uniformly convergent on  $S_0$ .

(3) *The Extended Schroedinger System for Degenerate States.*

Let  $\psi_{n0}^s$  ( $s=1, 2, \dots, t$ ) be  $t$  eigenfunctions belonging to the same eigenvalue  $K_{n0}$ . The B.C. of the  $\psi_{n0}^s$  are found from (8) or (10) respectively by replacing  $\psi_{n0}$  in turn by  $\psi_{n0}^s$  ( $s=1, 2, \dots$ ). We now follow the usual practice and introduce a new set of eigenfunctions  $\psi_{n0}'^s$  obtained from the  $\psi_{n0}^s$  by means of a unitary transformation with matrix  $\alpha$  and components  $(r|\alpha|s)$ ;

$$\psi_{n0}'^s = \sum_{r=1}^t (r|\alpha|s) \psi_{n0}^r. \quad (28)$$

As is easily seen the  $\psi_{n0}'^s$  functions will be orthonormal and have the same eigenvalues as the  $\psi_{n0}^s$ . Orthogonality follows from the fact that

$$\alpha \bar{\alpha}' = 1. \quad (29)$$

Instead of (14) we obtain from (13) by multiplication with  $\bar{\psi}_{n0}'^r$  and integration over  $D_0$

$$\int_{S_0} \left( \bar{\psi}_{n0}'^r \frac{\partial \psi_{n1}^s}{\partial n_0} - \frac{\partial \bar{\psi}_{n0}'^r}{\partial n_0} \psi_{n1}^s \right) dS = - \int_{D_0} \bar{\psi}_{n0}'^r (K_{n1} + H^1) \psi_{n0}^s d\tau. \quad (30)$$

We now substitute the expansion of  $\bar{\psi}_{n0}'^r$  from (28) into this equation and obtain

$$\sum_{l=1}^{\alpha} (r|\alpha|l) \{ (l|T|s) + (l|H^1|s) + (l|K|s) \} = 0, \quad (31)$$

$$(s=1, 2, \dots, t),$$

where  $r$  is supposed to remain fixed. The matrix elements used are defined as follows

$$(r|T|s) = \int_{S_0} \left( \bar{\psi}_{n0}'^r \frac{\partial \psi_{n1}^s}{\partial n_0} - \frac{\partial \bar{\psi}_{n0}'^r}{\partial n_0} \psi_{n1}^s \right) dS,$$

$$(r|H^1|s) = \int_{D_0} \bar{\psi}_{n0}'^r H^1 \psi_{n0}^s d\tau,$$

$$(r|K|s) = K_{n1} \delta_{rs}.$$

The set of equations obtained from (31) by varying  $s$  and keeping  $r$  fixed will be consistent only if the determinant of the  $t$  linear equations in the  $t$  unknowns  $(r|\alpha|t)$  vanishes, so that

$$|(r|T|s) + (r|H^1|s) + (r|K|s)| = 0. \quad (32)$$

Equation (32) gives us  $t$  roots for  $K_{n1}$ . For each one of these,  $K_{n1}^{\lambda}$  say, we obtain the associated set of  $(\lambda|\alpha|t)$  from (31), which fixes, however, only their ratios. Condition (29) determines their moduli uniquely. Equation (28) will then provide the appropriate  $\bar{\psi}_{n0}'^{\lambda}$ . For higher approximations one proceeds analogously.



In conclusion, I should like to express my grateful indebtedness to Prof. T. G. Cowling, of the University College of North Wales, for many valuable discussions, as well as critical and helpful suggestions.

*Bibliography.*

- L. Brillouin, *Comptes Rendus*, cciv. p. 1863 (1937).  
 N. Cabrera, *Comptes Rendus*, ccvii. p. 1175 (1938).  
 Feshbach and Clogstone, *Phys. Rev.*, lix. p. 189 (1941).  
 H. Feshbach, *Phys. Rev.*, lxv. p. 307 (1944).  
 H. Froehlich, *Phys. Rev.*, liv. p. 945 (1938).  
 K. Fuchs, *Proc. Roy. Soc. A*, clxxvi. p. 214 (1941).  
 G. Temple, *Proc. Roy. Soc. A*, clxix. p. 476 (1938).

LVII. *A Kinetic Energy Correction for the Flow of Plastic Liquids through Circular Pipes.*

By T. R. LOMER, B.Sc.\*

[Received May 6, 1946.]

RECENT advances in the field of theoretical plasticity have enabled the properties of plastic liquids to be measured in capillary instruments. The full investigation of such properties requires a wide range of rates of flow, and at high streaming velocities a kinetic energy correction is sometimes necessary.

The Poiseuille equation for the flow of Newtonian liquids through circular tubes may be corrected for kinetic energy losses by subtracting a term,  $mpv^2/\pi^2a^4$ , from the pressure applied across the ends of the flow tube, where  $m$  is a numerical constant,  $\rho$  the density of the liquid,  $v$  the volume flowing per second, and  $a$  the radius of the tube.

Couette (1890) found theoretically that  $m$  was unity; Boussinesq (1891) gave a more rigorous treatment and found  $m$  to be 1.12. The experimental determination of  $m$  is complicated by the difficulty of separating kinetic energy effects from "end correction" effects; according to Bingham <sup>(1)</sup>, if  $m$  is taken as 1.12 no end correction is necessary, whereas if  $m$  is taken as unity an end correction must be added to the length of the flow tube.

Schofield and Scott-Blair <sup>(2)</sup> have derived a general equation relating the rate of flow of a plastic liquid through a circular pipe of radius  $a$  and length  $l$ , to the shearing stress  $\tau_a (=Pa/2l)$  at its wall, the equation being

$$V = \frac{2\pi a^3}{\tau_a^3} \int_0^{\tau_a} \tau \int_{\tau}^{\tau_a} D(\tau) d\tau d\tau, \quad \dots \dots \dots (1)$$

where the shearing stress  $\tau$  at a distance  $r$  from the axis of the tube is given by

$$\tau = Pr/2l, \quad \dots \dots \dots (2)$$

and  $D(\tau)$  is a function characteristic of the particular liquid under consideration, and is defined as

$$dv/dr = -D(\tau), \quad \dots \dots \dots (3)$$

\* Communicated by the Author.

where  $dv/dr$  is the velocity gradient produced by the shearing stress  $\tau$ . For a Newtonian liquid  $dv/dr = -\tau/\eta$  and so  $D(\tau) = \tau/\eta$ . The kinetic energy correction term may be calculated in the general case as below. Let  $P$  be the total pressure difference across the ends of the tube,  $p$  be the pressure used in overcoming plastic resistance and  $W$  the kinetic energy imparted per second to the liquid. If  $V$  be the volume flowing per second through the tube, then

$$pV = PV - W. \quad (4)$$

Hence  $W/V$  is the correction term to be subtracted from the applied pressure. It follows from equations (2) and (3) that the velocity of flow at distance  $r$  from the axis of the tube is

$$v = \frac{a}{\tau_a} \int_{\tau}^{\tau_a} D(\tau) d\tau. \quad (5)$$

The volume of liquid flowing per second through a cylindrical shell of radius  $r$  is given by

$$dV = v \cdot 2\pi r dr.$$

Substituting for  $v$  from equation (5), and for  $r$  from equation (2), and integrating, we obtain (1).

Assuming now that the streaming velocity is developed in the flow tube and no appreciable fraction of it in the reservoir at the entrance to the tube, the kinetic energy  $dW$  produced per second in an elementary shell is  $dW = 2\pi r dr \rho v^3/2$ .

Making the same substitutions for  $v$  and  $r$  as previously, and integrating, we obtain

$$W = \pi \rho \left( \frac{a}{\tau_a} \right)^5 \int_0^{\tau_a} \tau \left( \int_{\tau}^{\tau_a} D(\tau) d\tau \right)^3 d\tau. \quad (6)$$

Hence

$$\frac{W}{V} = \frac{\rho}{2} \left( \frac{a}{\tau_a} \right)^2 \frac{\int_0^{\tau_a} \tau \left( \int_{\tau}^{\tau_a} D(\tau) d\tau \right)^3 d\tau}{\left( \int_0^{\tau_a} \tau \int_{\tau}^{\tau_a} D(\tau) d\tau \cdot d\tau \right)}.$$

This may be written as

$$\frac{W}{V} = \frac{K \rho V^2}{\pi^2 a^4}, \quad (7)$$

where

$$K = \frac{\tau_a^4 \int_0^{\tau_a} \tau \left( \int_{\tau}^{\tau_a} D(\tau) d\tau \right)^3 d\tau}{8 \left( \int_0^{\tau_a} \tau \int_{\tau}^{\tau_a} D(\tau) d\tau \right)^3}. \quad (8)$$

$K$  is dimensionless and on integration gives a numerical value (cf.  $m$  in the case of Newtonian liquids) which depends only on the relationship assumed between  $dv/dr$  and  $\tau$ . Because it has been assumed that no kinetic energy is produced outside the flow tube,  $K$  reduces to unity for Newtonian liquids in accordance with the value given by Couette; for liquids of the Ostwald type, which obey the law  $dv/dr = A\tau^B$  where

A and B are constants, K is given by

$$K = 3(B+3)^2 / 2(B+2)(3B+5). \quad \dots \quad (9)$$

As B approaches the value  $+\infty$ , then for  $\tau < 1$  the value of  $dv/dr$  approaches zero and K approaches the value 0.5. This corresponds to Bernoulli's

Fig. 1.

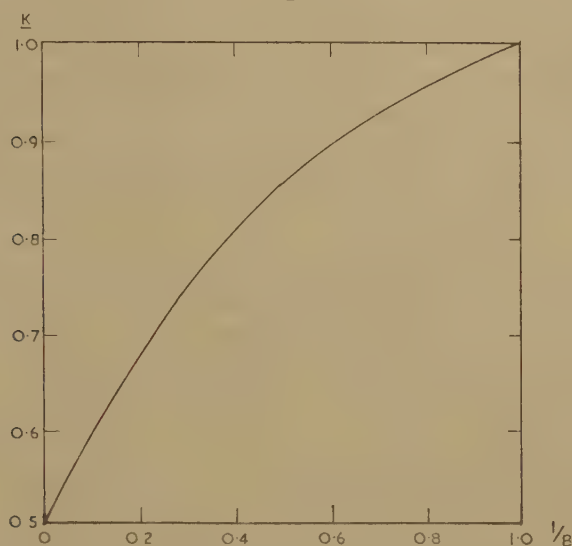
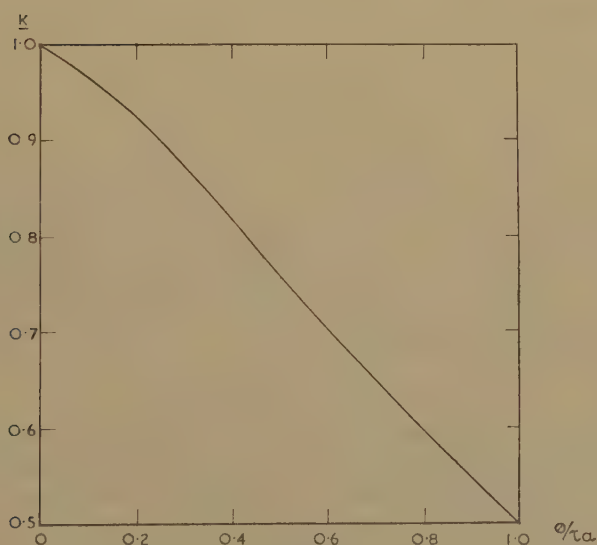


Fig. 2.



assumption that all the particles flow through the tube with the same velocity, and it is interesting to note that in these circumstances the correction is exactly one half of that given by Couette for a true liquid. When evaluating the correction term for liquids for which  $D(\tau)$  is not a

continuous function, care must be taken to split up the integrals in equation (8) and to integrate separately for each region in which  $D(\tau)$  is continuous. A Bingham liquid for which  $D(\tau)=0$  for  $0<\tau<\theta$  and  $D(\tau)=k(\tau-\theta)$  for  $\theta<\tau<\tau_a$  where  $\theta$  is the Bingham "yield value," gives

$$K=(1+58x/35+47x^2/35)/(1+2x/3+x^2/3)^3,$$

where  $x=\theta/\tau_a$ .

It should be noted that for  $\theta=0$ , *i. e.*  $x=0$ ,  $K=1$ , *i. e.* the value for a Newtonian liquid, and for  $\theta=\tau_a$  *i. e.*  $x=1$ , which corresponds to pure plug flow,  $K$  becomes equal to one half and again agrees with the value obtained on Bernoulli's assumption that all particles flow through the tube with the same velocity.

To facilitate the practical application,  $K$  is given as a graphical function of  $1/B$  for Ostwald-type liquids in fig. 1, and of  $\theta/\tau_a$  for Bingham liquids in fig. 2.

In conclusion I wish to express my thanks to the Directors of Lever Brothers and Unilever, Limited, for permission to publish this note.

#### References.

- (1) Bingham, 'Plasticity and Fluidity,' p. 29 (New York, 1922).
- (2) Scott Blair, 'Survey of General and Applied Rheology,' p. 136, (London, 1944).

44 Allport Road, Bromborough,  
Cheshire.

### LVIII. Partitioning of Space into Enantiomorphous Polyhedra.

By SIDNEY MELMORE\*.

[Received July 3, 1945.]

An enantiomorphous partitioning was described by Barlow in 1923 †. He arrived at it by considering a set of points under the operations of a certain space-group, and determining the shape and dimensions of the cell-walls bounding the region of operations about one of the points. This process involved laborious calculation. A partitioning of this character can, however, be obtained very simply, and its space-group be easily determined in the following way.

#### Construction of the Cell.

Fig. 1 depicts a rhombic dodecahedron with one of its hexagonal zone-axes set vertically. Consider any two alternate faces in this zone, say the left- and right-front faces in the figure. By "pointing" each of these faces, that is to say, by producing the surrounding faces until they meet in a common vertex, the non-self-reflexible polyhedron shown in fig. 2 is formed. Such a polyhedron and its enantiomorph, when repeated,

\* Communicated by the Author.

† W. Barlow, Phil. Mag. xlv. p. 930 (1923).



fill space without interstices. For the solid angle at each of the two new vertices are equal to one another and to the angle subtended at the centre of a rhombic dodecahedron by one of its faces : hence the dihedral angles about these vertices fulfil the requirements of the case ; and so also do the remaining dihedral angles of the polyhedron, for they are those of the original rhombic dodecahedron.

The polyhedron possesses a digonal axis and no other element of symmetry. It has ten faces. Taking the edges of the original rhombic dodecahedron as of unit length, the longer face-diagonal is 1.63 units : the faces of the derived polyhedron may then be described as follows :— There are three faces of the rhombic dodecahedron : six trapezia with edges 1.00, 1.00, 1.15, 2.00 and a short diagonal 1.63 : and a parallelogram with edges 1.15, 2.00 and a short diagonal 1.63. A net is shown in fig. 3.

#### *Determination of the Space-Group.*

If we take plastic spheres of equal size, bring them together according to Barlow's closest packed cubic arrangement \*, and then compress them

Fig. 1.

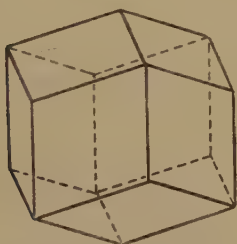
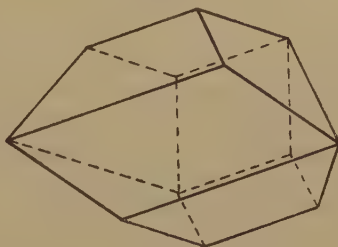


Fig. 2.

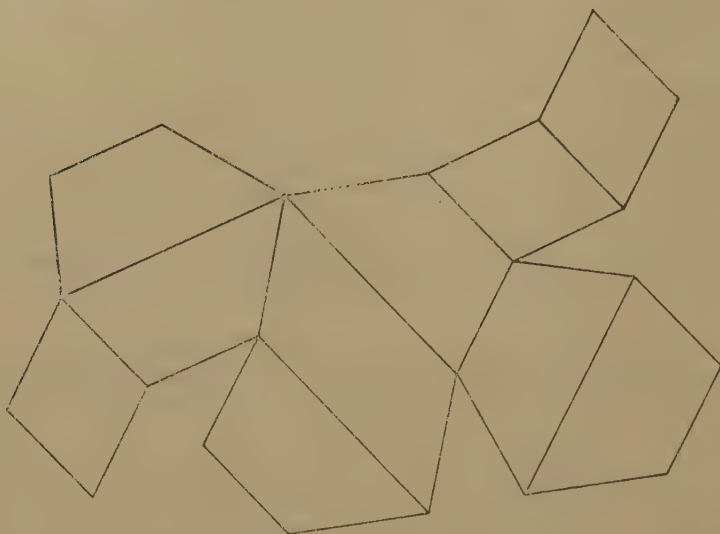


uniformly, the spheres become deformed into rhombic dodecahedra. Reverting now to the aggregate of enantiomorphous polyhedra it is evident, from the method of derivation, that it may be regarded as a collection of rhombic dodecahedra in which some are omitted, the place of each omitted one being represented by the region in which twelve new solid angles meet at a common vertex. After noting the position of these regions a model of rigid spheres can be built to represent the structure. Let the omitted dodecahedra be represented by red spheres and the others by green ones. Each sheet consists of a triangular arrangement of spheres in contact arranged as follows :—Make a row of spheres, one red to the left followed by three green ones to the right. Below, and to the right of the red sphere, begin another row, three green and one red. Begin a third row below and to the left of the second, consisting of one red and three green spheres. The centres of the spheres thus lie on the sides of two parallelograms (one the mirror-image of the other), the red spheres being at the extremities of the longer diagonals. The sheet is developed by translation of this set of spheres, a red in the new position coinciding with a red one in the old. The sheets are arranged with respect to one another as in Barlow's cubic structure and so that

\* W. Barlow, 'Nature,' xxix. p. 186 (1883).

each red sphere makes contact with three green ones in the sheet below and with three in the sheet above.

Fig. 3.



The model being completed, it will be seen at once that the red spheres lie at the corners of a rhombohedral lattice and that the content of the cell has the symmetry of the space-group  $R\bar{3}$ .

#### LIX. *Notices respecting New Books.*

*Relaxation Methods in Theoretical Physics.* By R. V. SOUTHWELL. [Pp. vi+248.] (Oxford University Press, 1946). Price 20s.

A STUDY of the present book, which is a continuation of the author's earlier treatise (1940), will convince the reader of the power of the relaxation method in solving a wide variety of problems arising in technical physics. The method can often produce results where an analytical method fails, and those concerned with the application of physics to practical problems, including research and design, will find it of great value to have a knowledge of the numerical methods described. A valuable feature of the book is the detailed treatment of twenty-four examples, taken from elasticity, hydrodynamics, electricity and magnetism, etc.

The type of problem which may be handled by the relaxation method is one governed by an equation of the form

$$\frac{\partial}{\partial x} \left( \chi(x, y) \frac{\partial \psi}{\partial x} \right) + \frac{\partial}{\partial y} \left( \chi(x, y) \frac{\partial \psi}{\partial y} \right) + Z(x, y) = 0,$$

where  $\chi$  and  $Z$  are specified functions. This equation is firstly approximated by a finite-difference equation, and for the benefit of readers not acquainted with the calculus of finite differences Chapter I. is devoted to relevant aspects of this. Chapters II. and III. are concerned with the details of the method in the case of problems satisfying Laplace's equation ( $\chi=1$ ,  $Z \equiv 0$ ) or Poisson's equation ( $\chi=1$ ). Several of the examples given are torsion problems in which either  $\psi$  or its normal gradient,  $\partial\psi/\partial n$ , takes specified values at the boundary.

But mixed boundary conditions, in which  $\psi$  is specified at some parts of the boundary and  $\partial\psi/\partial n$  at others, may also be handled and the torsion problem for an elliptical shaft illustrates this. Another type of boundary condition is "refraction" at an interface, and an example shows how the lines of magnetic force may be obtained in a field containing iron.

Chapter IV. is given to the important application of relaxation methods to conformal transformations. The shapes of boundaries encountered in many practical problems are inconvenient for direct application of the relaxation method, and this may often be overcome by a preliminary conformal-mapping which does not alter the type of equation but changes the boundary so that a convenient net may be cast for carrying out the relaxation technique. An interesting detailed example deals with the conformal transformation of a rail section into a rectangle. The method may also be applied to the computation of the electrical capacity of cables and condensers.

Problems governed by the general equation, where  $\chi$  is a function of  $x$  and  $y$ , are considered in Chapter V. These include the torsion of a circular non-uniform shaft, the temperature-distribution in the piston of an internal-combustion engine, and some problems in film lubrication. The great power of the relaxation method is demonstrated in Chapter VI., which deals with problems in which (1) different equations govern  $\psi$  in different parts of the field, or (2) some part of the boundary is not initially known. An example of (2), which is treated in detail, is the percolation of water through a retaining wall of porous material such as soil or peat.

Formulae and numerical tables for use in applying relaxation methods are given at the end of the book. The usual high standard of printing by the Clarendon Press, combined with the author's skill in exposition, make the book very attractive.

L. S. GODDARD.

*Advances in Colloid Science.* Edited by H. MARK and G. S. WHITBY.—Vol. II. *Scientific Progress in the Field of Rubber and Synthetic Elastomers.* [Pp. xl. + 453.] (New York: Interscience Publishers Inc., 1946.)

THIS volume includes a portrait and biography of the late Elmer O. Kraemer, nine monographs written by a total of twelve authors, a communal index, and an introduction dealing with work in fields connected with rubber problems. The style is often highly technical. Each monograph has its own introduction, summary and bibliography. The sum total of references is over 750.

Several of the authors have something to say, know what they are saying, and say it with appropriate objectivity. If books are to be "co-operative," however a large responsibility rests with the Editor (or Editors), whose duty it is to sift the material. In this case, a fair amount of chaff is mixed with the wheat, which is a pity, because the wheat is good, although some of it is unripe.

To those needing assistance in the further study of synthetic rubbers and elastomers generally, as well as of natural rubbers, the book is recommended as likely to provide a summary of the present position of the subject and to point the way to further information in the literature

J. R. P.

*Physical Methods of Organic Chemistry.* Edited by A. WEISSBERGER. Vol. II. [Pp. vii. + 737–1367.] (New York: Interscience Publishers Inc., 1946. Price £8.50.)

THE first volume of this work was reviewed in the 'Philosophical Magazine' in 1945 (vol. xxxvi. p. 658). The present volume deals with spectroscopy and spectrophotometry, colorimetry, polarimetry, dipole moments, conductometry, potentiometry, polarography, magnetic susceptibility, radioactivity, and mass spectrometry, and contains an index of both volumes. The book is intended to provide information to users of physical methods in organic chemistry, and is not an elementary work on general aspects of the physical side of organic chemistry suitable for students, other works of varying merit being available



on this side. The articles, therefore, are more concerned with the physical principles underlying the use of certain apparatus, and with the operation of this apparatus in the laboratory, than with the results which have been obtained with it. In this quite definite field, the book is on the whole very informative and helpful. Most of the modern methods are described, and there are good references to the literature. Apart from their practical value, many of the articles give a concise statement of the theory, that on polarimetry in the present volume, for example, containing a very lucid account of the general phenomena of polarized light. The book is well illustrated. The section on dipole moments is rather brief, and only one method is described in any detail. In view of its increasing importance, the section on electrophoresis is too short. The section on polarography, on the other hand, seems too detailed in comparison with the rest of the book. The general impression gained by the reviewer is that, apart from the section on polarimetry, this volume is less successful than the first, but the book, as a whole, is one which every chemical and physical laboratory should possess.

J. R. P.

*Beams on Elastic Foundation.* By M. HETÉNYI. [Pp. ix+255.] (The University of Michigan Press: Ann Arbor, 1946.) Price \$4.50.

THE problem of the elastically supported beam arises in numerous practical questions in civil and mechanical engineering; and in the present book a comprehensive analysis of the problem is made. The solution, in a large number of special cases, is given in a form which may be readily applied to the problem at hand. In Chapter I. the equation for the deflection ( $y$ ) of an elastic line, namely,

$$\frac{d^4y}{dx^4} + ky = q,$$

is established, and the solution in the case of no distributed load ( $q=0$ ) is given explicitly. Formulæ are obtained for the slope, moment and the shearing force. Chapter II. consists of an application of the results to infinite and semi-infinite beams, with various types of loading, and examples given include the railroad track and the deformation of a cylindrical tube under the action of radial forces acting at the points of a circle. Chapters III. and IV. are given to various problems involving finite beams, where the solutions may often be expressed in the form of trigonometric series. Examples studied include problems concerning an aqueduct, the commutator of an electrical machine, and the construction of a gangway in a coal seam.

When the beam has a variable flexural rigidity or modulus of foundation, the deflection satisfies a more complicated equation, but the solution is often expressible in terms of Bessel functions. This is explained in Chapter V. The problem of axial and transverse loading is analysed in Chapter VI. and the elastic stability of straight bars in Chapter VII. The next two chapters deal with the torsion of bars and with the problem of the circular arch. Application is made to the circular ring, the spherical shell and to the flexibility of corrugated tubes.

The last chapter deals with a different type of elastic foundation, namely, one where the deflection at any point depends on the deflections at other points (e.g., an elastic solid). The mathematics is still tractable, although more complicated. A good review is made of the literature in this case.

Tables of various functions are given at the end of the book, to facilitate application of the analytical solutions. The book should prove valuable to those concerned with the types of problems described. It provides easy reading, and it could not be said that the subject-matter is too condensed. On the contrary, there appears to be some unnecessary elaboration in a few places.

L. S. GODDARD.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



LX. *Further Notes and Suggestions on the Teaching of Physics.*

By C. J. SMITH, Ph.D., M.Sc., Lecturer in Physics, Royal Holloway College (University of London), Englefield Green, Surrey \*.

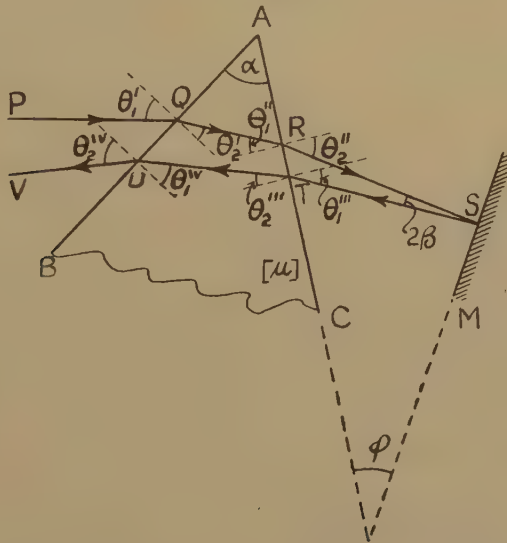
[Received May 31, 1945.]

PART II.†

*Angular Dispersion in a Littrow Spectrograph.*

IN the well-known Littrow spectrograph the ray of light which suffers minimum deviation in its passage through the prism is caused to fall normally on a plane mirror so that this ray retraces its path through the system. To calculate the angular dispersion in such an instrument, let  $\hat{BAC}$ , fig. 15, be the refracting angle,  $\alpha$ , of the prism and consider the

Fig. 15.



Angular dispersion in a Littrow spectrograph.

passage through the prism of a ray PQRS which falls at S on a plane mirror M and is reflected from there along the path STUV, the RST being, say,  $2\beta$ . Let the angles of incidence and refraction be as indicated. Then if  $\mu$  is the refractive index of the material of the prism for light of wave-length  $\lambda$ , we have, in the usual way,

$$\begin{aligned} \sin \theta_1' &= \mu \sin \theta_2', & \mu \sin \theta_1'' &= \sin \theta_2'', & \delta \theta_2' + \delta \theta_1'' &= 0, \\ \sin \theta_1''' &= \mu \sin \theta_2''', & \mu \sin \theta_1^{IV} &= \sin \theta_2^{IV}, & \delta \theta_2''' + \delta \theta_1^{IV} &= 0. \end{aligned}$$

\* Communicated by the Author.

† For Part I. c.f. Phil. Mag. Aug. 1946, p. 505.

Also

$$\theta_1''' + 2\beta = \theta_2'', \text{ and } \theta_1''' + \beta = \phi,$$

where  $\phi$  is the angle indicated—a constant for a given setting of the instrument. From these last two equations we have  $\delta\theta_1''' + \delta\theta_2'' = 0$ .

Differentiating, so that the changes in the above angles when the wave-length is  $\lambda + \delta\lambda$  may be obtained, we have, if the incident light is not dispersed,

$$\begin{aligned} 0 &= \mu \cos \theta_2' \delta\theta_2' + \delta\mu \sin \theta_2', \\ \mu \cos \theta_1'' \delta\theta_1'' + \delta\mu \sin \theta_1'' &= \cos \theta_2'' \delta\theta_2'', \\ \cos \theta_1''' \delta\theta_1''' &= \mu \cos \theta_2''' \delta\theta_2''' + \delta\mu \sin \theta_2''', \\ \mu \cos \theta_1^{\text{IV}} \delta\theta_1^{\text{IV}} + \delta\mu \sin \theta_1^{\text{IV}} &= \cos \theta_2^{\text{IV}} \delta\theta_2^{\text{IV}}. \\ \therefore \cos \theta_2^{\text{IV}} \delta\theta_2^{\text{IV}} &= \delta\mu \sin \theta_1^{\text{IV}} - \cos \theta_1^{\text{IV}} \mu \delta\theta_2''' \\ &= \delta\mu \sin \theta_1^{\text{IV}} - \cos \theta_1^{\text{IV}} [(\cos \theta_1''' \delta\theta_1''' \\ &\quad - \delta\mu \sin \theta_2''') / \cos \theta_2'''] \\ &= (\delta\mu \sin \alpha + \cos \theta_1^{\text{IV}} \cos \theta_1''' \delta\theta_2'') / \cos \theta_2''' \\ &= \frac{\delta\mu \sin \alpha}{\cos \theta_2'''} + \frac{\cos \theta_1^{\text{IV}} \cos \theta_1'''}{\cos \theta_2''' \cos \theta_2''} [\delta\mu \sin \theta_1'' \\ &\quad + \mu \cos \theta_1'' (-\delta\theta_2')] \\ &= \frac{\delta\mu \sin \alpha}{\cos \theta_2'''} + \frac{\cos \theta_1^{\text{IV}} \cos \theta_1'''}{\cos \theta_2''' \cos \theta_2''} \left[ \delta\mu \sin \theta_1'' \right. \\ &\quad \left. + \frac{\cos \theta_1'' \sin \theta_2'}{\cos \theta_2'} \delta\mu \right] \\ &= \frac{\delta\mu \sin \alpha}{\cos \theta_2'''} + \frac{\cos \theta_1^{\text{IV}} \cos \theta_1'''}{\cos \theta_2''' \cos \theta_2''} \left[ \frac{\delta\mu \sin \alpha}{\cos \theta_2'} \right]. \end{aligned}$$

For the position of minimum deviation this gives, if  $\psi$  is then the angle of emergence, *i. e.* at the position of minimum deviation  $\theta_2^{\text{IV}} = \psi$ ,

$$\cos \psi \frac{\partial \psi}{\partial \lambda} = \frac{\partial \mu}{\partial \lambda} \left[ \frac{\sin \alpha}{\cos \frac{1}{2}\alpha} \right] + \frac{\cos \frac{1}{2}\alpha \cos \psi}{\cos \frac{1}{2}\alpha \cos \psi} \left[ \frac{\sin \alpha}{\cos \frac{1}{2}\alpha} \frac{\partial \mu}{\partial \lambda} \right],$$

*i. e.* the angular dispersion  $\frac{\partial \psi}{\partial \lambda}$  is given by

$$\frac{\partial \psi}{\partial \lambda} = \frac{4 \sin \frac{1}{2}\alpha}{\cos \psi} \frac{\partial \mu}{\partial \lambda}.$$

Since  $\frac{\partial \mu}{\partial \lambda}$  is negative, it follows that the rays of wave-length  $\lambda + \delta\lambda$  will give rise to a spectral line on the L.H.S. of the line  $\lambda$  when an observer looks on to the focusing screen.

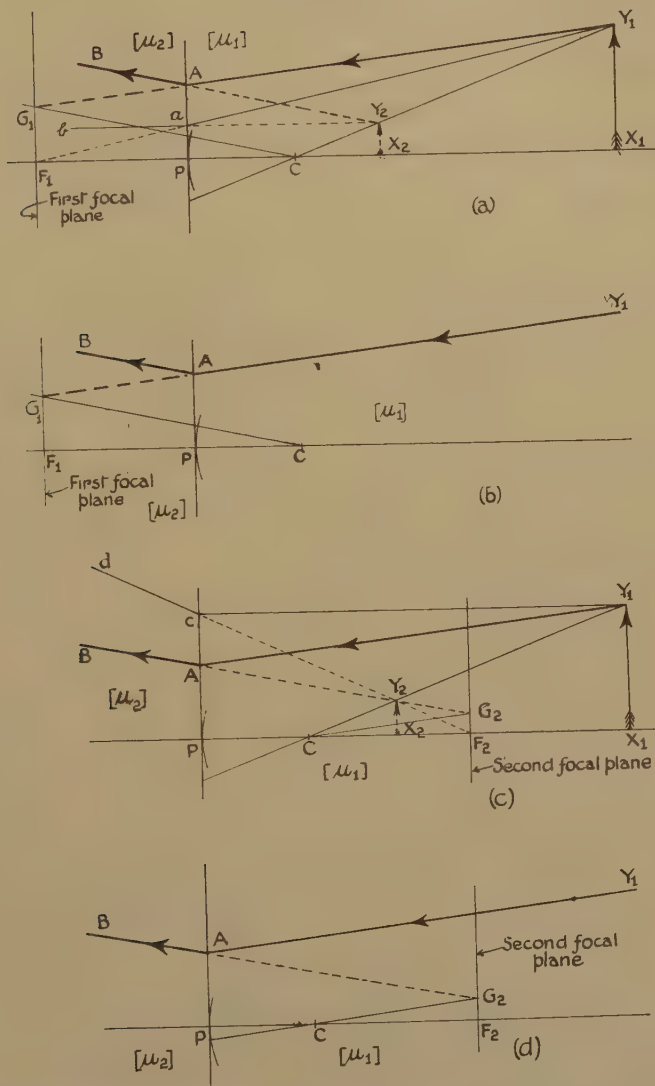
#### *On a General Method of Tracing Paraxial Rays refracted at a Spherical Interface: Some Important Properties of the Focal Planes.*

In teaching the fundamental principles of geometrical optics, not sufficient attention is paid always to some important properties of focal planes and, even when these planes are mentioned, no attempt is made to establish these properties theoretically by a simple argument. Such an argument is as follows.

Our starting point, the object being a point source of light, is the

usually accepted definition of the first and second principal focal points, viz. "The limiting positions of the object and the image when the image and object are respectively at infinity are known as the first and second principal foci of the given optical system." Planes through these points

Fig. 16.



Refraction at a spherical interface : fundamental properties of the focal planes and the tracing of paraxial rays.

[N.B.—All distances normal to the principal axis are highly magnified in the diagrams.]

and normal to the principal axis are termed the first and second focal planes. Now let P, C and  $F_1$ , fig. 16 (a), be the pole, centre of curvature and first principal focus, when a spherical surface of radius  $r$  separates

two media of absolute refractive indices  $\mu_1$  and  $\mu_2$ . It will be assumed that  $\mu_1 < \mu_2$ . Let it be required to trace the path of a paraxial ray  $Y_1A$  after refraction at the interface. First consider a somewhat laborious construction which, however, is necessary in order to establish theoretically the method finally used. Let  $X_1Y_1$  be a small finite object normal to the principal axis of the system. To determine the position of the image of  $X_1Y_1$  formed by refraction at the interface we use a ray  $Y_1a$ , directed towards  $F_1$ , which after refraction at  $a$  travels along  $ab$  in a direction parallel to  $CP$ , and also the ray  $Y_1C$  which is not deviated. If the directions of the two refracted rays intersect in  $Y_2$ , then  $Y_2$  is the image of  $Y_1$  and the complete image may be obtained in the usual way. If  $Y_2$  is joined to  $A$  by a straight line and this is produced to  $B$ ,  $AB$  is the required refracted ray.

Now suppose that  $Y_1A$ , produced, cuts the first focal plane in  $G_1$ . The following analysis shows that  $CG_1$  is parallel to  $AB$  and it is this fact which finally enables us considerably to simplify the method of constructing the path of the refracted ray. In the proof which follows all distances will be considered numerically. If  $AY_2$  is parallel to  $CG_1$ , then

$$(G_1A)/(AY_1) = (CY_2)/(Y_2Y_1).$$

Since  $F_1G_1$  is parallel to  $X_1Y_1$ ,

$$(G_1A)/(AY_1) = (F_1P)/(PX_1) = [(\mu_1|r|)/(\mu_2 - \mu_1)]/|x_1| = (\mu_1 R)/\{(\mu_2 - \mu_1)X_1\},$$

if  $R$ ,  $X_1$ , and  $X_2$  denote the magnitudes of  $r$ ,  $x_1$  and  $x_2$  respectively.

Also since  $X_2Y_2$  is parallel to  $X_1Y_1$ ,

$$(CY_2)/(Y_2Y_1) = (CX_2)/(X_2X_1) = (|x_2| - |r|)/(|x_1| - |x_2|) = (X_2 - R)/(X_1 - X_2).$$

Now from the fundamental paraxial equation

$$\mu_1[(1/r) - (1/x_1)] = \mu_2[(1/r) - (1/x_2)],$$

we have for the system under discussion

$$\mu_1[(1/R) - (1/X_1)] = \mu_2[(1/R) - (1/X_2)],$$

from which it follows that

$$X_2 = (\mu_2 R X_1) / \{(\mu_2 - \mu_1)X_1 + \mu_1 R\}.$$

Hence 
$$(CY_2)/(Y_2Y_1) = [\{(\mu_2 R X_1)/(\mu_2 - \mu_1 X_1 + \mu_1 R)\} - R] \div [X_1 - (\mu_2 R X_1)/\{\mu_2 - \mu_1 X_1 + \mu_1 R\}]$$

$$= (\mu_1 R)/(\mu_2 - \mu_1 X_1).$$

Thus the theorem is established. Its physical significance lies in the fact that it proves that if a converging pencil of light is directed towards a point  $G_1$ , say, in the first focal plane of a refracting interface, then the emergent pencil is parallel to  $CG_1$ . The more elegant method of determining the direction of the emergent ray corresponding to  $Y_1A$  is shown in fig. 16 (b).  $Y_1A$  is produced to cut the first focal plane in  $G_1$ , and this point is joined to  $C$  by a straight line. The refracted ray  $AB$  is parallel to  $CG_1$ .

In the above, use has been made of the first principal focus and the first focal plane. A similar method may be adopted in which use is made



of the second focal plane. Thus, in fig. 16 (c) the refracted image is located by using (i) a ray  $Y_1C$  parallel to  $CP$ , which is refracted so that it appears to pass through  $F_2$ , and (ii) the ray  $Y_1C$  which is not deviated. The refracted ray,  $AB$ , is found as in the earlier method. When making use of the first focal plane we let the incident ray  $Y_1A$  cut that plane in  $G_1$ : now let the refracted ray  $AB$  cut the second focal plane in  $G_2$ . We have to show that  $CG_2$  is parallel to  $Y_1A$ . If these two straight lines are parallel, then

$$(CY_2)/(Y_2Y_1) = (G_2Y_2)/(Y_2A),$$

all distances being considered numerically as before. Now

$$(CY_2)/(Y_2Y_1) = (CX_2)/(X_2X_1) = (X_2 - R)/(X_1 - X_2) = (\mu_1 R)/\{(\mu_2 - \mu_1)X_1\},$$

$$\text{and } (G_2Y_2)/(Y_2A) = (X_2F_2)/PX_2 = (PF_2 - |x_2|)/|x_2|$$

$$= \{(\mu_2 R)/(\mu_2 - \mu_1) - X_2\}/X_2 = (\mu_1 R)/\{(\mu_2 - \mu_1)X_1\}.$$

Thus the theorem is established. Its importance physically lies in the fact that if one ray of a parallel pencil falling on a spherical interface cuts the second focal plane in  $G_2$ , then all the refracted rays corresponding to those in the incident pencil appear to come from  $G_2$ . A variation of the method of constructing the path of the emergent ray when the incident ray is  $Y_1A$  is therefore as follows. In fig. 16 (d) a straight line  $G_2C$  is drawn parallel to the incident ray  $Y_1A$ . If  $G_2C$  cuts the second focal plane in  $G_2$ , then  $G_2A$  produced gives the required refracted ray  $AB$ .

An exactly similar method may be used to establish the relevant theorem in the case of reflecting surfaces.

To examine the corresponding problem in the case of a lens system comprising two thin lenses arranged coaxially in air, let  $O_1$  and  $O_2$ , fig. 17 (a), be the centres of the thin lenses, the cardinal points being as shown.

[In order to obtain an accurate diagram the cardinal points have been calculated for the system  $f_1 = -15$  cm.,  $f_2 = 30$  cm. and  $a = 10$  cm., the focal length of a converging lens being considered a negative quantity.]

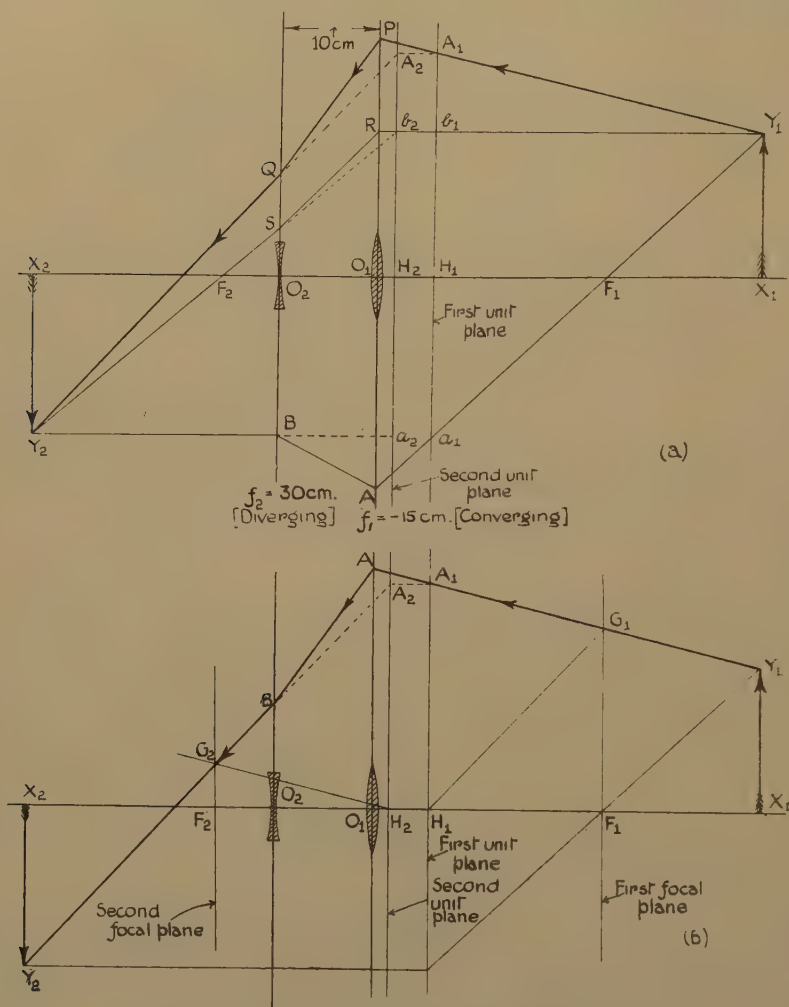
As before, let  $X_1Y_1$  be a small object normal to the principal axis of the system. To locate the final image formed by refraction through the system,  $Y_1F_1$  is drawn to cut the first unit plane in  $a_1$ , when the corresponding refracted ray, parallel to the principal axis of the system, is  $a_2Y_2$ , where  $H_1a_1 = H_2a_2$ , and then  $Y_1b_1$  is drawn to cut the second unit plane in  $b_2$  when the corresponding emergent ray is along  $b_2F_2$ , where  $H_1b_1 = H_2b_2$ . As a check use may be made of the fact that  $Y_1H_1$  is parallel to  $H_2Y_2$ . To find the actual path between the lenses of any of the rays used in locating  $X_2Y_2$  we note, for example, that if  $Y_1a_1$  cuts the first lens in  $A$  and  $Y_2a_2$  cuts the second lens in  $B$ , then  $Y_1ABY_2$  is the complete path of the ray selected from  $Y_1$  to  $Y_2$ . Similarly for  $Y_1b_1b_2RSY_2$  and other rays.

To trace the path of any incident ray when  $X_2Y_2$  has been located we note that a ray  $Y_1P$  cuts the first unit plane in  $A_1$  and after refraction through the system it travels along  $A_2Y_2$ , where  $H_1A_1 = H_2A_2$ . If, moreover,  $Y_1P$  cuts the first lens in  $P$ , while  $A_2Y_2$  cuts the second lens

in  $Q$ , then  $Y_1PQY_2$  is the complete path of the ray we are considering from  $Y_1$  to  $Y_2$ .

To discover and justify a more elegant method for tracing the path of any ray through the system, let us suppose that the incident ray intersects the first focal plane in  $G_1$ : then—cf. fig. 17 (b)—if  $G_1H_1$  is parallel

Fig. 17.



Ray tracing through a coaxial system of lenses.

to the emergent ray  $A_2Y_2$  the new method is at once apparent. To establish the fact that these lines are parallel let us imagine that the left-hand portion of the diagram is moved so that  $H_2$  coincides with  $H_1$  and  $A_2$  with  $A_1$ . Then, since  $H_1A_1$  and  $F_1G_1$  are each parallel to  $X_1Y_1$ , we have, all distances being considered numerically,

$$(A_1G_1)/(G_1Y_1) = (H_1F_1)/(F_1X_1) = (|f|)/(|\xi_1| - |f|). \quad [\xi_1 = H_1X_1].$$

Similarly,

$$\begin{aligned} (Y_2H_2)/(H_1Y_1) &= (X_2H_2)/(H_1X_1) = \{\xi_2\}/\{\xi_1\}, & [\xi_2 = H_2X_2] \\ &= \{f\}/\{\xi_1 - |f|\}, & [\because 1/|\xi_2| + 1/|\xi_1| = 1/|f|] \\ \therefore (A_1G_1)/(G_1Y_1) &= (Y_2H_2)/(H_1Y_1). \end{aligned}$$

Hence  $G_1H_1$  is parallel to  $A_2Y_2$ . Thus, to construct the path of the refracted ray when the positions of the cardinal points of the system are known, let  $G_1$  be the point in which the incident ray cuts the first focal plane of the system. Join  $H_1G_1$ . If the first unit plane is cut by the incident ray in  $A_1$ , we make  $H_2A_2 = H_1A_1$  and draw through  $A_2$  a straight line parallel to  $G_1H_1$ . This straight line gives the path of the emergent ray when  $Y_1A$  is the incident ray.

[In a similar way it may be shown that if a straight line is drawn through  $H_2$  parallel to the incident ray, to cut the second focal plane in  $G_2$ , then  $A_2G_2$  gives the path of the emergent ray.]

When the positions of the cardinal points of a thick lens are known the paths of the rays through it are obtained by methods identical with those used above for tracing paraxial rays through a coaxial lens system. The properties of the focal planes of a thick lens are identical with those of a coaxial lens system : no separate proofs will therefore be given.

#### *Image Formation by Refraction through a Solid Sphere of Isotropic Material.*

When an image of a point object is produced by refraction at a spherical interface the fundamental paraxial relation becomes

$$\mu_1/\{(\rho/\zeta_1) - 1\} = \mu_2/\{(\rho/\zeta_2) - 1\},$$

where  $\rho$  is the radius of curvature measured from the centre of curvature of the spherical surface, and  $\zeta_1$  and  $\zeta_2$  are the object and image distances measured from the same point. The above equation is more easily remembered if it is written in the form

$$\mu_1[(1/\rho) - (1/\zeta_2)] = \mu_2[(1/\rho) - (1/\zeta_1)],$$

or, better,  $(1/\mu_1)[(1/\rho) - (1/\zeta_1)] = (1/\mu_2)[(1/\rho) - (1/\zeta_2)],$

since the form of the equation is now more symmetrical : in fact, it may be said that  $(1/\mu)[(1/\rho) - (1/\zeta)]$  is an optical invariant.

As a check we may put  $\mu_1 = -\mu_2$ , when the corresponding formula in the case of a spherical mirror should be obtained. It is, of course,  $(1/\zeta_1) + (1/\zeta_2) = (2/\rho)$ , which is easily established directly.

The advantages sometimes gained by measuring distances from the common centre of curvature of several refracting surfaces are revealed by considering a particular example :—

EXAMPLE.—A glass sphere, diameter 4 cm., whose material has a refractive index  $3/2$ , is surrounded on one side by air and on the other by a liquid whose refractive index is  $4/3$ . Determine the positions of the cardinal points. Stops limit the rays to the paraxial region.

Let  $X_1$  be the position of a point object,  $X'$  that of the image formed by refraction at the air-glass interface, while  $X_2$  is the position of the final image. In fig. 18 (a) and (b) the positions of these points are shown, first with respect to  $C$ , the centre of curvature of both interfaces, and secondly with respect to  $O_1$  and  $O_2$ , the vertices of the interfaces. Various distances are represented by the symbols shown on the diagram. At the head of the work which follows there is given the fundamental paraxial equation in its appropriate form for the two methods of solving the problem.

$$(1/\mu_1)[(1/\rho) - (1/\zeta_1)] = (1/\mu_2)[(1/\rho) - (1/\zeta_2)]$$

For the refraction at the first surface, we have

$$1[(1/2) - (1/\zeta_1)] = (2/3)[(1/2) - (1/\zeta')], \quad (i)$$

while at the second

$$2/3[-(1/2) - (1/\zeta')] = 3/4[-(1/2) - (1/\zeta_2)]. \quad (ii)$$

The position of  $F_1$  is obtained by making  $\zeta_2 \rightarrow \infty$ . Equation (ii) gives  $[\zeta']_{\zeta_2 \rightarrow \infty} = 16$  cm.

Substituting this value in (i) we get

$$(\zeta_1)_{\zeta_2 \rightarrow \infty} = 4.8 \text{ cm.}$$

Now make  $\zeta_1 \rightarrow \infty$  in order to locate  $F_2$ . Equation (i) gives  $[\zeta']_{\zeta_1 \rightarrow \infty} = -4$  cm., so that from (ii) we have

$$[\zeta_2]_{\zeta_1 \rightarrow \infty} = -3.6 \text{ cm.}$$

The nodal points coincide at the centre of the sphere. To determine the unit points a general expression is required for  $m$ , the lateral magnification. If we imagine the system to rotate through a small angle about  $C$ , so that  $X_1$  moves to  $Y_1$  and  $X_2$  to  $Y_2$ , then from the similar  $\Delta$ 's  $X_1CY_1$  and  $X_2CY_2$ , we have

$$m = \zeta_2/\zeta_1.$$

Equations (i) and (ii) may be written

$$(3/2)[(1/2) - (1/\zeta_1)] = (1/2) - (1/\zeta'),$$

$$\text{and } 9/8[-(1/2) - (1/\zeta_2)] = -(1/2) - 1/\zeta'.$$

Eliminating  $\zeta'$ , we have

$$-(3/2)(1/\zeta_1) + (9/8)(1/\zeta_2) = -5/16, \quad (iii)$$

$$\therefore \zeta_2/\zeta_1 = (5/24) \zeta_2 + (3/4).$$

If  $m = \zeta_2/\zeta_1 = 1$ ,  $(1/4) = (5/24)[\zeta_2]_{m=1}$ ,

$$\therefore [\zeta_2]_{m=1} = 1.2 \text{ cm.}$$

Thus  $H_2$  is located. To locate  $H_1$ , without using the fact that  $H_1$  and  $H_2$  must coincide since  $N_1$  and  $N_2$  do, we have from (iii)

$$-(3/2) + (9/8)(\zeta_1/\zeta_2) = -(5/16) \zeta_1,$$

$$\therefore (1/m) = (\zeta_1/\zeta_2) = -(5/18) \zeta_1 + (4/3).$$

If  $m = 1$ ,  $-(1/3) = -(5/18)[\zeta_1]_{m=1}$ ,

$$\text{i. e., } [\zeta_1]_{m=1} = 1.2 \text{ cm.}$$

$$\mu_1[(1/r) - (1/x_1)] = \mu_2[(1/r) - (1/x_2)].$$

For the refraction at the first surface, we have

$$-(1/2) - (1/x_1) = (3/2)[-(1/2) - (1/x')], \quad (i)$$

while at the second

$$(3/2)[(1/2) - 1/(x' + 4)] = (4/3)[(1/2) - (1/x_2)]. \quad (ii)$$

These give, after some reduction,

$$x_1/x' = (4 - x_1)/6, \quad (iii)$$

$$\text{or } x' = (6x_1)/(4 - x_1). \quad (iv)$$

$$\text{Also } (x' + 4)/x_2 = (14 - x')/16. \quad (v)$$

The position of  $F_1$  is obtained by making  $x_2 \rightarrow \infty$ . Equation (v) gives  $[x']_{x_2 \rightarrow \infty} = 14$  cm.

Using this value of  $x'$  in (iv) we get

$$[x_1]_{x_2 \rightarrow \infty} = 2.8 \text{ cm.}$$

To locate  $F_2$ , let  $x_1 \rightarrow \infty$ . Equation (iv) gives

$$[x']_{x_1 \rightarrow \infty} = -6 \text{ cm.}$$

Using (v) we have

$$-2/[x_2]_{x_1 \rightarrow \infty} = 5/4.$$

$$\therefore [x_2]_{x_1 \rightarrow \infty} = -1.6 \text{ cm.}$$

To find the unit points, we have, if  $m$  is the lateral magnification,

$$\frac{1}{m} = \frac{(x' + 4)/(3/2)}{x_2/(4/3)} \cdot \frac{(x_1/1)}{x'/(3/2)}$$

$$= \frac{(4/3)(x' + 4)}{x_2} \cdot \frac{x_1}{x'}$$

$$= \frac{4}{3} \cdot \frac{14 - x'}{16} \cdot \frac{4 - x_1}{6}$$

$$= \frac{4}{3} \cdot \frac{14 - \{6x_1/(4 - x_1)\}}{16} \cdot \frac{4 - x_1}{6}$$

$$= (56 - 20x_1)/72.$$

$$\text{If } m = 1, \quad 72 = 56 - 20[x_1]_{m=1},$$

$$\therefore [x_1]_{m=1} = -0.8 \text{ cm.}$$

Substituting this value of  $x_1$  in (iv), we get

$$[x']_{m=1} = -1 \text{ cm.}$$

Hence from (v)

$$[x_2]_{m=1} = 3.2 \text{ cm.}$$



The cardinal points are therefore as shown in fig. 18 (c).

[It is easily verified that

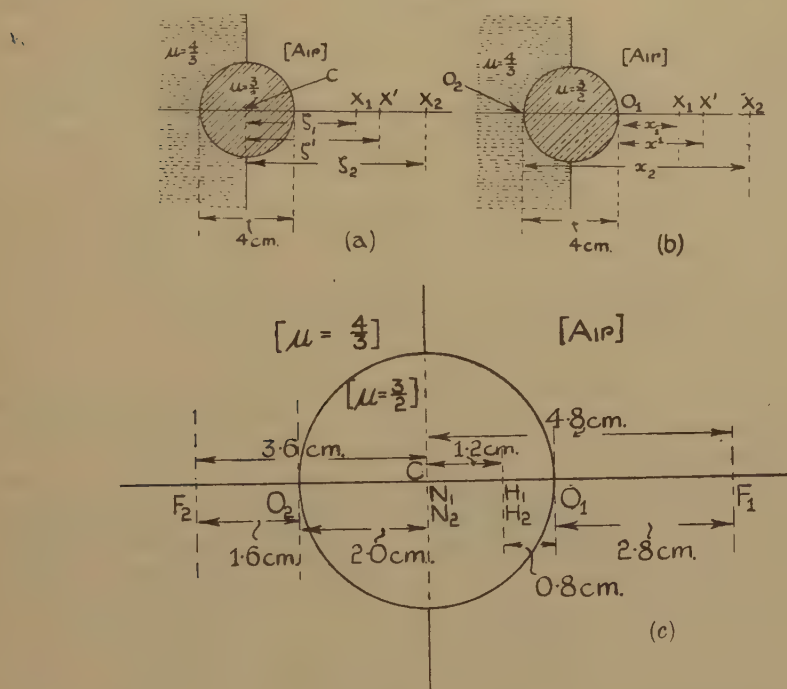
$$(H_1 F_1)/1 + (H_2 F_2)/(4/3) = 0,$$

$$N_1 F_1 + H_2 F_2 = 0,$$

$$N_2 F_2 + H_1 F_1 = 0,$$

as is required by the general theory of systems of coaxial refracting surfaces.]

Fig. 18.



### Experimental Determination of the Positions of the Cardinal Points of a Thick Lens when the Initial and Final Media are different.

Let  $r_1$  and  $r_2$  be the radii of curvature of the faces of a thick lens, the refractive index of whose material is  $\mu$  and whose thickness is  $t$ , surrounded on one side by a medium of refractive index  $\mu_1$  and on the other by a medium of refractive index  $\mu_2$ . Let  $X_1$  be the position of a point object on the principal axis of the system, while  $X$  and  $X_2$  are its successive images formed in turn by the refraction of paraxial rays at the surfaces of the lens. Then if  $x_1$  and  $x$  are the distances of  $X_1$  and  $X$  from  $O_1$ , while  $x_2$  is the distance  $O_2 X_2$ , where  $O_1$  and  $O_2$  are the vertices of the lens, then the appropriate paraxial equations are

$$\mu/x - \mu_1/x_1 = (\mu - \mu_1)/r_1 = \kappa_1, \quad \text{and} \quad \mu_2/x_2 - \mu/(x+t) = (\mu_2 - \mu)/r_2 = \kappa_2,$$

where  $\kappa_1$  and  $\kappa_2$  are the refracting powers of the surfaces of the lens.

Let  $\sigma$  denote a "reduced distance" i. e.  $\sigma = x/\mu$ . Let  $\tau = t/\mu$ . Then the above formulæ become

$$1/\sigma - 1/\sigma_1 = \kappa_1 \quad \text{and} \quad 1/\sigma_2 - 1/(\sigma + \tau) = \kappa_2,$$

so that  $\sigma_2 = (\alpha\sigma_1 + \beta)/(\gamma\sigma_1 + \delta)$ , where  $\alpha = \tau\kappa_1 + 1$ ,  $\beta = \tau$ ,  $\gamma = \tau\kappa_1\kappa_2 + \kappa_1 + \kappa_2$ ,  $\delta = \tau\kappa_2 + 1$ .

Let  $f_1$  be the first focal distance. Then

$$f_1/\mu_1 = [x_1/\mu_1]_{x_1 \rightarrow \infty} = [\sigma_1]_{\sigma_1 \rightarrow \infty} = -\delta/\gamma,$$

$$\therefore f_1 = -\mu_1(\delta/\gamma).$$

Similarly

$$f_2 = \mu_2(\alpha/\gamma).$$

It then follows almost at once that for any pair of conjugate points  $X_1$  and  $X_2$ ,  $F_1X_1 \cdot F_2X_2 = -(\mu_1\mu_2)/\gamma^2$ .

The positions of the unit points  $H_1$  and  $H_2$  are determined from the relations  $1/m = \gamma\sigma_1 + \delta$  and  $m = \alpha - \gamma\sigma_2$ , where  $m$  is the lateral magnification. We have

$$O_1H_1 = \mu_1(1 - \delta)/\gamma; \quad O_2H_2 = \mu_2(\alpha - 1)/\gamma.$$

Combining these equations, we get  $H_1F_1 = -\mu_1/\gamma = \phi_1$ , say, and  $H_2F_2 = \mu_2/\gamma = \phi_2$ , say, so that  $\phi_1/\mu_1 + \phi_2/\mu_2 = 0$ .

If  $\chi_1$  and  $\chi_2$  denote  $F_1X_1$  and  $F_2X_2$  respectively, we have

$$F_1X_1 \cdot F_2X_2 = \phi_1\phi_2, \quad \text{or} \quad \chi_1\chi_2 = \phi_1\phi_2,$$

which is the general form of Newton's equation, almost invariably written  $pq = -f^2$  when applied to a single lens or to a coaxial lens system, but in the general form the absence of the negative sign should be noted.

To find the positions of the nodal points,  $N_1$  and  $N_2$ , let  $y_1$ ,  $y$  and  $y_2$  denote the sizes of a small object normal to the axis of the system and its successive images. Then, at the nodal points, Helmholtz's equation becomes  $\mu_1y_1 = \mu_2y_2$ . But  $y_2/y_1$  is the linear magnification, so that  $y_2/y_1 = 1/(\gamma\sigma_1 + \delta) = \alpha - \gamma\sigma_2$ . Hence

$$O_1N_1 = (\mu_2 - \mu_1\delta)/\gamma, \quad \text{and} \quad O_2N_2 = (\mu_2\alpha - \mu_1)/\gamma.$$

From these we have,

$$H_1F_1 + N_2F_2 = 0; \quad N_1F_1 + H_2F_2 = 0; \quad \text{and} \quad H_1H_2 = N_1N_2.$$

Now let  $\xi_1$  and  $\xi_2$  be the distances  $H_1X_1$  and  $H_2X_2$ , respectively. Then from fig. 19 (a), where the position of the image  $X_1Y_1$  is determined graphically, in a particular case, we have  $y_1/y_2 = (\xi_1 - \phi_1)/(-\phi_1)$ , and  $y_1/y_2 = (-\phi_2)/(\xi_2 - \phi_2)$ , from which we get  $(\xi_1 - \phi_1)(\xi_2 - \phi_2) = \phi_1\phi_2$ , or  $\frac{\phi_1}{\xi_1} + \frac{\phi_2}{\xi_2} = 1$ . Since  $\phi_1/\mu_1 + \phi_2/\mu_2 = 0$ , we have  $1/\phi_2 = 1/\xi_2 - (\mu_1/\mu_2)(1/\xi_1)$ , which is most easily remembered if it is written in the form

$$\frac{1}{(\xi_2/\mu_2)} - \frac{1}{(\xi_1/\mu_1)} = \frac{1}{(\phi_2/\mu_2)} = -\frac{1}{(\phi_1/\mu_1)}.$$

The reason why the unit points are also called the principal points is at once revealed.

Now  $m$  is  $(\xi_2/\mu_2) \div (\xi_1/\mu_1)$  so that from the above equation we have at once

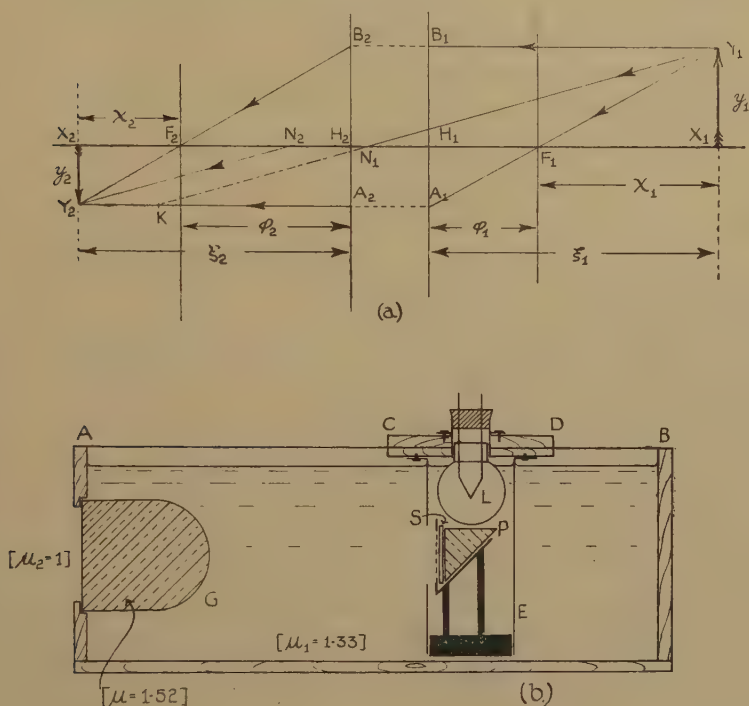
$$1 - m = \xi_2/\phi_2.$$

In an experiment  $\xi_2$  cannot be measured unless the position of  $H_2$  is known. Let the position of the image be measured from any convenient point. If the above distance is  $\xi_2'$ , we have  $\xi_2 = \xi_2' + A$ , where  $A$  is a constant. Thus  $1 - m = (\xi_2' + A)/\phi_2$ , so that if corresponding values of  $m$  and of  $\xi_2'$  are plotted,  $\phi_2$  may be determined from the slope of the straight line so obtained.

Similarly,  $(1/m) - 1 = -(\xi_1)/\phi_1$ , so that  $\phi_1$  may be found.

A convenient piece of apparatus for locating the positions of the cardinal points is shown in fig. 19 (b).

Fig. 19.



The cardinal points of a thick lens when the initial and final media are different.

AB is a wooden box, 4 ft. by 4 in. by 4 in., into one end of which a thick glass lens  $G$  is fixed. A convenient lens consists of a glass cube surmounted by a hemisphere, the axis being about 7 cm. long. The object consists of two thin parallel wires supported in front of a prism  $P$  and illuminated by a 12 volt, 36 watt lamp  $L$ , the box being filled with water. A white translucent screen,  $S$ , is placed between the wires and the prism. The wires and prism, etc., are fixed to a wooden board which slides along the top of the box  $AB$ . A scale in cm., etc., is fixed to one edge of the box. A bituminous compound may be used to make the box watertight. To determine the position of the first principal focus,  $F_1$ , a telescope,

focused for parallel light, is directed along the axis of the system and the slide carrying P, etc., adjusted until a clear image of the wires is seen in the telescope. To locate  $F_2$  illuminated wires are placed on the L.H.S. of the system and a plane mirror in the liquid and facing G used in the ordinary way.

To determine  $m$  an image of the wires illuminated by L is obtained on a ground glass screen and the size of the image determined either with a pair of dividers or, better, with a low power travelling microscope : in the last instance the screen should be removed. Values for  $\phi_1$  and  $\phi_2$  may then be obtained in the manner suggested above.

The apparatus does not locate the positions of the nodal points but these may be determined graphically when the positions of  $H_1$ ,  $H_2$ ,  $F_1$  and  $F_2$  are known. Suppose they are as in fig. 19 (a). Let the position of an image  $X_2Y_2$  of an object  $X_1Y_1$  be found graphically. Then if  $Y_2K$  is set off along  $Y_2A_2$  so that  $Y_2K = H_2H_1$ , then  $Y_1K$  passes through  $N_1$  :  $N_2$  is easily found for  $H_1H_2 = N_1N_2$ .

### *Oblique Refraction of a Thin Pencil at a Single Spherical Surface.*

On proceeding to discuss the refraction of a thin pencil of light at a single spherical surface and then extending the discussion to a symmetrical optical instrument, it is assumed that the chief ray of the pencil and the principal axis are in a plane but that the chief ray is inclined to the axis at an angle which is not small. Whatever may be the number of refractions which subsequently occur, the chief ray never leaves the above plane, which is termed the *meridional plane*, of the pencil of light. From considerations of symmetry the two principal sections of the pencil are those made by the meridional plane and a plane at right angles to this and passing through the chief ray. These two sections are known respectively as the *meridional* (or *primary*) *section* and the *sagittal* (or *secondary*) *section* of the pencil. The corresponding points to which rays in the meridional and the sagittal planes are brought to a focus are known as the *meridional* (or *primary*) *focus* and the *sagittal* (or *secondary*) *focus* of the pencil : they are respectively the centres of curvature of the meridional and sagittal sections of the wave-front.

Let C, fig. 20 (a), be the centre of curvature of a spherical surface separating two media of absolute refractive indices  $\mu_1$  and  $\mu_2$  respectively. Let  $P_1A$  be the chief ray of a thin pencil whose meridional focus is  $P_1$ . Let  $AP_2$  be the refracted ray. Consider  $P_1B$ , a ray adjacent to  $P_1A$  ; after refraction, let it traverse the path  $BP_2$ , so that when A and B become very close together  $P_2$  assumes a definite limiting position, viz., the meridional focus of the refracted pencil. Let  $\theta_1$  and  $\theta_2$  be the angles of incidence and refraction for the chief ray : for the adjacent rays let the corresponding angles be  $\theta_1 + \delta\theta_1$  and  $\theta_2 + \delta\theta_2$ . Then

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2,$$

whence

$$\mu_1 \cos \theta_1 \delta\theta_1 = \mu_2 \cos \theta_2 \delta\theta_2.$$



Let  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\epsilon$  be the small angles indicated. Then

$$\theta_1 + \delta\theta_1 = \gamma + \beta = \alpha + \theta_1 + \beta, \quad \text{i. e., } \delta\theta_1 = \alpha + \beta,$$

and  $\theta_2 + \delta\theta_2 + \epsilon = \beta + \theta_2, \quad \text{i. e., } \delta\theta_2 = \beta - \epsilon.$

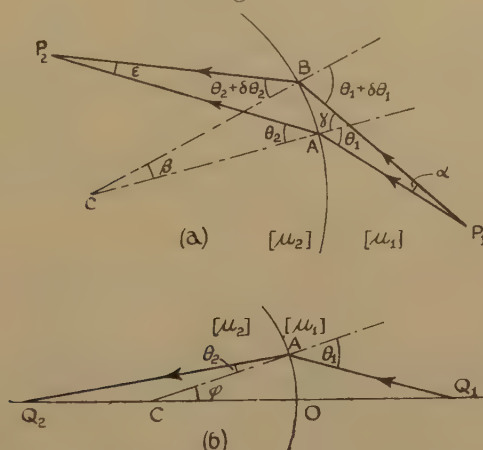
We therefore have

$$\begin{aligned} \mu_1 \cos \theta_1 \cdot (\alpha + \beta) &= \mu_2 \cos \theta_2 \cdot (\beta - \epsilon), \\ \therefore \mu_1 \cos \theta_1 [(AB \cos \theta_1)/|p_1| + (AB)/|r|] \\ &= \mu_2 \cos \theta_2 [(AB)/|r| - (AB \cos \theta_2)/|p_2|], \end{aligned}$$

where  $p_1 = AP_1$ ,  $p_2 = AP_2$  and  $r$  is  $AC$ . Now  $p_1$  is positive, while  $p_2$  and  $r$  are negative. Hence

$$(\mu_1 \cos^2 \theta_1)/p_1 - (\mu_2 \cos^2 \theta_2)/p_2 = (\mu_1 \cos \theta_1 - \mu_2 \cos \theta_2)/r.$$

Fig. 20.



Oblique refraction at a spherical interface.

This is the equation connecting the positions of corresponding primary foci due to refraction at a single spherical interface. It may be written in the symbolic form

$$\Delta[(\mu \cos^2 \theta)/p] = (1/r) \Delta[\mu \cos \theta],$$

where, in general,

$$\Delta[z] = z_1 - z_2.$$

It is now necessary to determine the position of the sagittal focus. Let  $Q_1$ , fig. 20 (b), be a point source of light on the axis through C, the centre of curvature of the interface. By symmetry, the secondary focus of the refracted pencil must lie on  $CQ_1$  or on  $Q_1C$  produced. Let it be  $Q_2$ . Then the chief ray is refracted along  $AQ_2$ . Let  $\theta_1$ ,  $\theta_2$  and  $\phi$  be the angles indicated. Then Snell's law gives

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2,$$

$$\therefore \mu_1 (\sin \theta_1)/\sin \phi = \mu_2 (\sin \theta_2)/\sin \phi.$$

i. e. 
$$\mu_1 \cdot \frac{CQ_1}{AQ_1} = \mu_2 \cdot \frac{CQ_2}{AQ_2}.$$

Now

$$|CA|=|r|=|CQ_1| \cos \phi - |AQ_1| \cos \theta_1,$$

$$\therefore |CQ_1|=\{|r|+|AQ_1| \cos \theta_1\}/\cos \phi,$$

and

$$|r|=|AQ_2| \cos \theta_2 - |CQ_2| \cos \phi,$$

i. e.,

$$|CQ_2|=\{-|r|+|AQ_2| \cos \theta_2\}/\cos \phi.$$

$$\therefore \mu_1\{|r|+|AQ_1| \cos \theta_1\}/|AQ_1|=\mu_2\{-|r|+|AQ_2| \cos \theta_2\}/|AQ_2|.$$

$$\therefore \mu_1/|AQ_1|+\mu_2/|AQ_2|=(-\mu_1 \cos \theta_1+\mu_2 \cos \theta_2)/|r|.$$

$$\therefore (\mu_1/q_1)-(\mu_2/q_2)=(\mu_1 \cos \theta_1-\mu_2 \cos \theta_2)/r,$$

with the usual sign convention, if  $q_1$  and  $q_2$  denote the distances of  $Q_1$  and  $Q_2$  from A respectively. In symbolic notation this becomes

$$\Delta[\mu/q]=(1/r)\Delta[\mu \cos \theta].$$

When  $\theta_1 \rightarrow 0$  and  $\theta_2 \rightarrow 0$  the above equations each reduce to

$$\Delta[\mu/x]=(1/r)\Delta[\mu],$$

which is the ordinary paraxial equation for refraction at a spherical interface.

The positions of the focal lines formed by reflexion at a spherical surface may be obtained at once by substituting  $\mu_1=-\mu_2$  in the formulæ just obtained. They are given by

$$(1/p_1)+(1/p_2)=2/(r \cos \theta),$$

and

$$(1/q_1)+(1/q_2)=(2 \cos \theta)/r.$$

To test the last two formulæ experimentally we may use the arrangement described by Searle\*, although the author prefers to use a transparent cross illuminated with monochromatic light as the object. In such an experiment  $p_1$  is necessarily equal to  $q_1$ , so that we have

$$(1/p_1)+(1/p_2)=2/(r \cos \theta), \quad \text{and} \quad (1/p_1)+(1/q_2)=(2 \cos \theta)/r.$$

Eliminating  $p_1$ , we get

$$(1/p_2)=(1/q_2)+2[\{1/(r \cos \theta)\}-\{\cos \theta/r\}].$$

If, therefore, we keep  $\theta$  constant and vary  $p_1$  to obtain corresponding values of  $p_2$  and  $q_2$ , then by plotting  $y=1/p_2$  and  $x=1/q_2$ , we should obtain a straight line whose slope is  $+1$  and whose intercept on the  $y$ -axis is  $(2 \sin^2 \theta)/(r \cos \theta)$ .

#### *On the Theory and Use of a Jamin Interferometer.*

The starting point in the theory of this instrument is the assumption that  $(\mu-1)$ , the refractivity of the gas, is directly proportional to  $\rho$ , its density. If  $\mu_0$  and  $\rho_0$  refer to S.T.P. we have, in so far as the gas may be considered an ideal one,

$$\frac{\mu-1}{\mu_0-1}=\frac{\rho}{\rho_0}=\frac{T_0}{T} \cdot \frac{p}{p_0},$$

---

\* Searle, 'Experimental Optics,' C. U. P., p. 243.

where  $T$  is the temperature of the gas on the absolute scale, whence

$$\delta\mu = [(\mu_0 - 1)/p_0](T_0/T)\delta p,$$

if the temperature of the gas is kept constant.

If the tube containing the gas under investigation is of length  $l$  and  $\lambda$  is the wave-length of the monochromatic light used [ $\lambda$  is measured in the gas at pressure  $p$  and temperature  $T$ ], the number of light waves in the column of gas of length  $l$  is  $N = l/\lambda = (\mu l)/\lambda_0$ , where  $\lambda_0$  is the wave-length of the light in a vacuum, since  $c = f\lambda$ ,  $c_0 = f\lambda_0$ , where  $f$  is the frequency of the light and  $c$  its velocity, so that  $\mu\lambda = \lambda_0$ .

Hence, if the refractive index of the gas is changed (by altering the pressure)

$$\delta N = \delta\mu \cdot (l/\lambda_0) = \{(\mu_0 - 1)l/(\lambda_0 p_0)\}(T_0/T)\delta p,$$

or

$$\left(\frac{\partial N}{\partial p}\right)_T = \frac{(\mu_0 - 1)l}{\lambda_0 p_0} \cdot \frac{T_0}{T}.$$

If the pressure of the gas in the comparison tube remains constant,  $\delta N$  is equal to the number of fringes which cross the field of view when the pressure is changed by  $\delta p$ . If  $\Delta[N]$  denotes the number of fringes which pass when the pressure is increased from  $p_1$  to  $p_2$ , we have

$$\Delta[N] = \int_{p_1}^{p_2} dN = \frac{(\mu_0 - 1)l}{\lambda_0 p_0} \frac{T_0}{T} (p_2 - p_1) = \frac{(\mu_0 - 1)l}{\lambda_0 p_0} \cdot \frac{T_0}{T} P,$$

where  $P = p_2 - p_1$ , the increase in pressure. Thus if  $\Delta[N]$  is plotted against  $P$ , we should obtain a straight line from the slope of which  $(\mu_0 - 1)$  may be determined.

Now, strictly speaking, it is necessary to correct for the fact that the mercury in the pressure gauge is not at  $0^\circ \text{C}$ .—the correction for the expansion of the scale is negligible. We have  $P = g\sigma H$ , where  $\sigma$  is the density of mercury at the temperature of the gauge,  $g$  the intensity of gravity and  $H$  is the height in cm. of the mercury column corresponding to  $P$ . Also  $p_0 = g\sigma_0 (76)$ .

$$\therefore \Delta[N] = \frac{(\mu_0 - 1)l}{76 \lambda_0} \cdot \frac{T_0}{T} \cdot \frac{\sigma}{\sigma_0} \cdot H.$$

Since  $\sigma_{20} = 13.545 \text{ gm. cm.}^{-3}$  and  $\sigma_0 = 13.595 \text{ gm. cm.}^{-3}$ , the correction is almost 0.5 per cent., which is about the error in this experiment when carried out as a laboratory exercise.

### *On Gauss' Experimental Verification of the Inverse Square Law of Magnetism.*

If  $\theta_A$  and  $\theta_B$  are the deflections of a magnetometer needle when placed in turn at a distance  $r$  from the centre of a short bar magnet and in the tangent A and B positions of Gauss respectively, it is well-known that  $\tan \theta_A = n \tan \theta_B$ , if the law of force is an inverse  $n$ -th power one in so far as the distance between magnetic poles is concerned.

Now in the usual account given in text-books to-day it is frequently stated that Gauss\* used short magnets when he established experimentally

---

\* Gauss, *Ann. der Phys.* xxviii. p. 241 and p. 591 (1833).

that  $n$  was equal to 2. Actually he used a magnet which was a foot long and had a mass of about one pound. The magnetometer was of the reflecting type and carried a plane mirror attached rigidly to the small magnet. A horizontal scale in mm. etc., and more than a metre in length, was fixed above a telescope five metres away from the magnetometer needle. The telescope was arranged so that the image of the horizontal scale above it, which was formed by reflexion in the plane mirror, could be observed. A displacement of one scale division across the fiducial mark in the observing telescope corresponded to an angular deflexion of 22 seconds of arc. A rotation of one-tenth this amount could be estimated by eye. Now Gauss was fully appreciative of the fact that the chief uncertainty in these experiments is in estimating the distance between the poles of the magnet. He therefore decided to work with the magnet at distances between 1 and 4 metres from the magnetometer, so that the fraction  $1/r^3$  was always small and certainly negligible when  $r$  exceeded 2 metres.

Some of the results which Gauss obtained between the 24th and 28th June, 1832, are shown in the following table :—

Distance of magnet from magnetometer ( $r$ ).	Deflexion of needle.	
	Tan A position.	Tan B position.
1.1 metres	2° 13' 51".2	1° 10' 19".3
2.0    "	37' 16".2	19' 1".6
4.0    "	4' 35".9	2' 22".2

Since  $\tan \theta_A = 2 \tan \theta_B$ , it was proved that  $n$  was equal to 2 within the limits of experimental error.

It is interesting to note that with the same apparatus Gauss made the first absolute determination of  $H_0$ , the horizontal component of the earth's magnetic field. Until then, values of  $H_0$  at different places had only been compared by an oscillation method.

*On the Determination of the Constants in the Thermoelectric Equation  $E = \alpha\theta + \frac{1}{2}\beta\theta^2$ , and the Construction of Lead-iron Thermocouples.*

Experiment shows that the e.m.f. of a thermocouple, at least over a limited temperature range, may be represented by the equation  $E = \alpha\theta + \frac{1}{2}\beta\theta^2$ , when one junction is at the temperature of melting ice and  $\theta$  is the temperature of the other junction on a centigrade scale referred to a gas thermometer:  $\alpha$  and  $\beta$  are constants for a given thermocouple. Now the above equation may be written

$$E/\theta = \alpha + \frac{1}{2}\beta\theta,$$

which is of the form  $y = a + bx$ , so that  $\alpha$  and  $\beta$  may be obtained from the graph obtained by plotting  $(E/\theta)$  against  $\theta$ .

To construct a lead-iron junction, in order to determine the thermoelectric power of iron, is a matter of some difficulty: the author has



found the following method very successful and one easily learned by students:—Pure Swedish iron wire S.W.G. No. 30, and obtainable from Messrs. Johnson Matthey & Co. Ltd. London, is used. The end to be joined to a piece of lead is freed from grease by immersion for about one minute in the usual chromic acid cleaning mixture and then washed with distilled water. The clean end of the iron wire must then be held for about 20 seconds in a strong aqueous solution of copper sulphate so that it becomes coated with a thin layer of copper. This is then washed with water and smeared with a soldering flux—"Coraline" may be used. This prepared end is then dipped into molten lead and withdrawn—it is found to be coated with lead. The lead wire, S.W.G. No. 20, is then scraped and wrapped round the lead-covered iron wire, the joint-to-be being smeared with flux. The two wires are held in forceps at a point about 0.5 cm. away from the prepared junction, a small wad of wet cotton-wool between the wires and the forceps protecting the wires from mechanical damage and also serving to keep the lead as cool as possible in the last part of the process. This consists in plunging the junction into some lead which is just about to solidify and withdrawing the wires quickly. A joint prepared in this manner is "good" electrically.

#### *On the Variation of the Resistance of Metals with Temperature.*

By measuring the resistance of a wire at different temperatures it has been found that, in general, the resistance increases as the temperature is raised. For most pure metals the curve showing the variation of resistance with temperature is almost a perfect parabola and, for platinum, Callendar has shown that the variation of resistance with temperature may be very accurately represented by the equation  $R_\theta = R_0(1 + \alpha\theta + \beta\theta^2)$ , where  $\theta$  is the temperature on the centigrade scale of temperature with nitrogen as the working substance,  $R_\theta$  being the resistance at  $\theta^\circ\text{C}$ . and  $R_0$  that at  $0^\circ\text{C}$ . To determine the constants  $\alpha$  and  $\beta$  it is necessary to measure the resistance of a coil of platinum wire at three well-known temperatures, *e.g.*, those of melting ice, steam, and sulphur vapour. These are  $0^\circ\text{C}$ .,  $100^\circ\text{C}$ . and  $444.6^\circ\text{C}$ ., respectively, if, in the last two instances the barometer reads 76 cm. of mercury under standard conditions and the vapours are not superheated.

Now the above equation may be written  $\{(R_\theta - R_0)/R_0\}\theta^{-1} = \alpha + \beta\theta$ , so that  $\alpha$  and  $\beta$  may be calculated when the values of the L.H.S. of this equation are known at two temperatures other than  $0^\circ\text{C}$ . Of course, a graphical method could be used by making observations on  $R_\theta$  at several known temperatures.

#### *On Centigrade Resistance Scales of Temperature.*

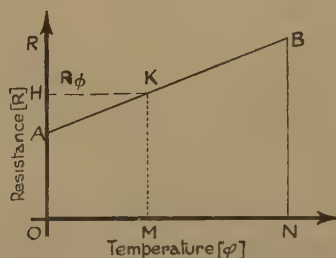
If  $R_0$  and  $R_{100}$  are the resistances of any given piece of wire at the temperatures of melting ice and of steam produced under standard conditions, a centigrade scale of temperature may be constructed by drawing two ordinates OA and BN, fig. 21, to represent to scale  $R_0$  and

$R_{100}$ , where the distance ON is 100 equal arbitrary units. The points A and B are joined by a straight line. To find the temperature  $\phi$  on this scale, corresponding to a resistance value  $R_\phi$ , OH is constructed to represent  $R_\phi$ , HK drawn parallel to ON to intersect AB in K, and KM drawn normal to ON to intersect this line in M. Then OM is  $\phi$  units long so that  $\phi$  becomes known.

For temperatures measured on this scale the relationship between  $R_\phi$  and  $R_0$  is a linear one and may be written  $R_\phi = R_0(1 + \kappa\phi)$ , where  $\kappa$  is a constant known as *the coefficient of increase of resistance with temperature as measured on the specified centigrade scale of temperature for the metal under investigation*.

Now it so happens that for pure metals  $\kappa$  is very nearly equal to  $\alpha$ , which occurs in the equation  $R_\theta = R_0(1 + \alpha\theta + \beta\theta^2)$ , which means that  $\beta$  is a small quantity: it is also found that  $\kappa \doteq \alpha \doteq (1/273) \text{ deg.}^{-1} \text{ C.}$  For alloys  $\kappa$  and  $\alpha$ , while still remaining approximately equal for any given alloy, vary considerably from one alloy to another and in the case of manganin, minalpha, and constantan tend to zero.

Fig. 21.



Construction of a centigrade scale of temperature.

The above argument seems to be a better approach to the subject than that generally given and brings out clearly the difference between temperatures measured on various centigrade scales of temperature. A moment's reflexion on the meaning of the equation  $l_\theta = l_0(1 + \lambda\theta)$ , which refers to the linear expansion of a metal, shows that strictly speaking it should be written  $l_\phi = l_0(1 + \kappa\phi)$ , where  $\phi$  is the temperature on a centigrade scale fixed with reference to the expansion of the given rod between the temperature of melting ice and that of steam produced under standard conditions. Similar remarks apply to equations representing the specific volume of a substance and its variation with temperature.

*Note on Laplace's Formula for the Magnetic Field due to an Element of Current: the Field due to a Linear Current.*

An exact statement of this fundamental law in electromagnetism seldom appears in English text-books and, as a result, much confusion of thought arises. Suppose that a steady current  $i$  (e.m.u.) is flowing in a wire: consider the magnetic field at a point P, fig. 22 (a), due to the current in that element AB of the wire, where, if O is a fixed point in the

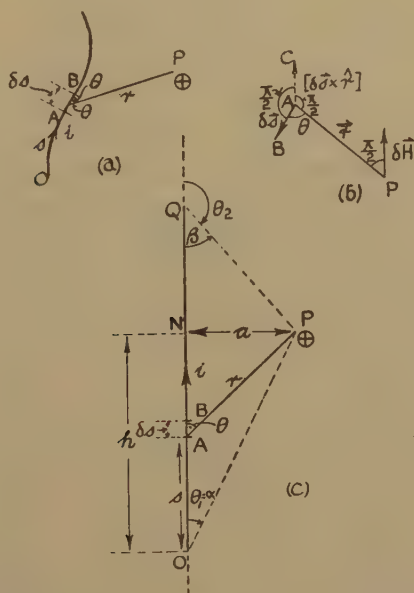
wire,  $OA=s$  and  $AB=\delta s$ . Let  $AP=r$  and let  $AB$  (not  $BA$ ) make an angle  $\theta$  with  $AP$ . Then if  $\delta H$  is the field strength at  $P$ , Laplace's formula is  $\delta H=(i \delta s \sin \theta)/r^2$ , and  $\delta H$  is directed downwards if  $P$  lies in the plane of the diagram. In vector notation we may write

$$\delta \vec{H}=i[\delta \vec{s} \times \hat{r}]/r^2=i[\delta \vec{s} \times \vec{r}]/r^3,$$

where  $\hat{r}$  is unit vector along  $AP$ . Now  $[\delta \vec{s} \times \hat{r}]$  is a vector normal to the plane  $PAB$  and of magnitude  $\sin \theta \cdot \delta s$ . The direction and sense of  $\delta \vec{H}$  are therefore as shown in fig. 22 (b).

Now consider the strength of the magnetic field at a point  $P$ , distance  $a$  from a straight wire  $OQ$ , fig. 22 (c), in which a steady current  $i$  is flowing.

Fig. 22.



Laplace's law and the magnetic field due to a linear current.

Let  $AB$  be an element of the wire, where  $OA=s$  and  $AB=\delta s$ . Then  $\delta H=i \delta s \sin \theta/r^2$ . Let  $h=ON$ , where  $N$  is the projection of  $P$  on the wire. Then  $s=h-a \cot \theta$ , or  $\delta s=a \operatorname{cosec}^2 \theta \delta \theta$ , and  $r=a \operatorname{cosec} \theta$ . Hence

$$H=\frac{i}{a} \int_{\theta_1}^{\theta_2} \sin \theta d\theta,$$

where  $\theta_1$  and  $\theta_2$  are the angles indicated. We have at once

$$H=(i/a)[\cos \theta_1-\cos \theta_2]=(i/a)[\cos \alpha+\cos \beta],$$

if we put  $\theta_1=\alpha$  and  $\pi-\theta_2=\beta$ .

### Experimental Determination of the Resistivity of Mercury.

This experiment is suitable for university classes and is important since it affords a means of demonstrating how to calibrate a capillary tube, which is an essential part of the apparatus. If  $\chi$  is the resistivity

of mercury at the temperature of the experiment, we have  $R = (\chi l) / (\pi r^2)$ , if  $l$  is the length of the tube and  $r$  its internal radius :  $R$  is the resistance of the column of mercury. In the above formula it is assumed that the bore of the tube is uniform. In practice this condition is never fulfilled, so that the tube must be calibrated : one method is as follows. The actual measurements are identical with those which have to be made in calibrating a tube for use in a viscometer \*. We have

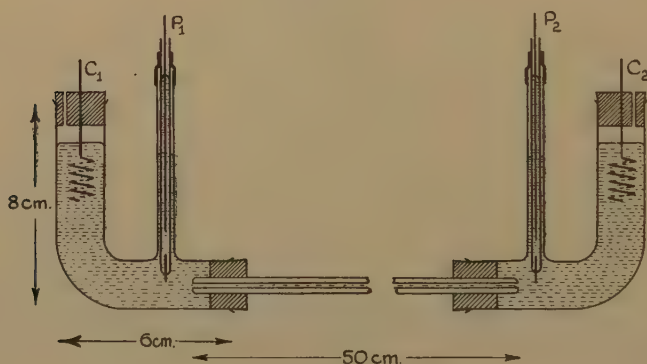
$$R = \frac{\chi}{\pi} \sum_{k=1}^n \frac{\Delta l_k}{r_k^2}.$$

Also, as before,  $\pi \rho r_k^2 \lambda_k = C$ , and  $M = \frac{l}{n} C \sum_{k=1}^n \left( \frac{1}{\lambda_k} \right)$ .

$$\begin{aligned} \therefore R &= \frac{\chi}{\pi} \frac{l}{n} \sum_{k=1}^n \left( \frac{1}{r_k^2} \right) = \frac{\chi}{\pi} \frac{l}{n} \frac{\pi \rho}{C} \sum_{k=1}^n (\lambda_k) \\ &= \chi \frac{l^2}{n^2} \frac{\rho}{M} \sum_{k=1}^n \left( \frac{1}{\lambda_k} \right) \cdot \sum_{k=1}^n (\lambda_k). \end{aligned}$$

Thus  $\chi$  may be determined when  $R$  has been measured. A suitable tube for this experiment is 50 cm. long and about 1 mm. in diameter. The ends of the tube are rounded slightly so that  $l$  may be measured accurately. The capillary tube is fitted into glass tubes of the shape and size indicated in fig. 23 :  $C_1$  and  $C_2$  are current leads ;  $P_1$  and  $P_2$  are

Fig. 23.



Resistivity of mercury.

potential leads. The potential probes consist of pyrex glass tubes into which short pieces of tungsten wire have been sealed. These tubes are partly filled with mercury and copper leads are used to make the necessary connections to a potentiometer. The current leads are made of iron, or nickel, so that no amalgams are produced. The resistance of the mercury column may then be measured in the usual way, the accuracy obtainable being limited mainly by the errors in the resistance coils and the

\* C. J. Smith, *Phil. Mag.* vol. xxxvii. p. 517, Aug. 1946.



temperature changes to which the apparatus is subjected, unless it is kept at a constant temperature such as that of melting ice. A correction must then be applied since the tube will not have been calibrated at this temperature.

*On the Comparison of Resistances from Observations on the Damping of a Moving Coil Galvanometer.*

If  $I$  is the moment of inertia of the suspended coil about its axis of rotation,  $\theta$  the angular deflection measured from the undisplaced position of the coil,  $a\dot{\theta}$  a frictional couple resisting the motion and  $b\theta$  the restoring couple due to the rigidity of the material of the suspension, then the equation of motion is

$$I\ddot{\theta} + a\dot{\theta} + b\theta = 0,$$

or

$$\ddot{\theta} + \alpha\dot{\theta} + \beta\theta = 0,$$

where  $I\alpha = a$  and  $I\beta = b$ .  $\alpha$  is termed the damping coefficient.

In addition to the mechanical damping represented by the term  $a\dot{\theta}$ , it has to be remembered that, if the coil forms part of a closed circuit, the movement of the coil in the field of the permanent magnet will induce currents which tend to bring the coil more quickly to rest. For simplicity, in investigating the extent of this additional couple, it will be assumed that the coil moves in a uniform magnetic field  $\vec{H}$ : in actual practice the magnetic field is part radial so that the magnetic flux through the coil varies as it swings, but the actual calculation of this change in flux would be difficult—cf. fig. 24 (a) and (b).

Let  $\Phi$  be the flux at any given instant, i. e.,  $\Phi = \int \vec{B} \cdot \hat{n} ds = \int \vec{H} \cdot \hat{n} ds$ , since the surrounding medium is air, and let  $N$  be the number of turns, each of area  $A$ , in the coil. Then  $AH \sin \theta = \Phi$  and the induced e.m.f. is given by

$$e = -N\dot{\Phi} = -ANH \cos \theta \cdot \dot{\theta}.$$

Thus the instantaneous current,  $i$ , in the coil is given by

$$i = -(ANH \cos \theta \cdot \dot{\theta})/r,$$

where  $r$  is the resistance of the complete circuit in e.m.u. The meaning of the negative sign is that if  $\theta$  is increasing, the direction of the current is opposite to that shown in fig. 24 (c). In this diagram the positive direction of the current is shown conventionally—it is such that the magnetic field due to it has a component in the same sense as  $\vec{H}$ .

Now the potential energy  $W$ , associated with the coil and due to the fact that it is carrying a current and is situated in a magnetic field, is given by

$$\begin{aligned} W &= -i \times \text{flux of magnetic induction} \times \text{the number of turns} \\ &= (A^2 N^2 H^2 \sin \theta \cos \theta \cdot \dot{\theta})/r. \end{aligned}$$

$\therefore$  Couple acting on the coil due to the induced current in it and tending to increase  $\theta$ , is

$$-\frac{\partial W}{\partial \theta} = -(A^2 N^2 H^2 / r) \left[ \{ \cos^2 \theta - \sin^2 \theta \} \dot{\theta} + \sin \theta \cos \theta \frac{\partial \dot{\theta}}{\partial \theta} \right].$$

When  $\theta \rightarrow 0$ , the above couple reduces to  $-[(A^2 N^2 H^2)]/\dot{\theta}$ . The equation of motion is therefore

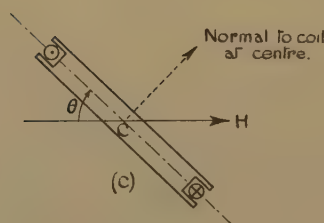
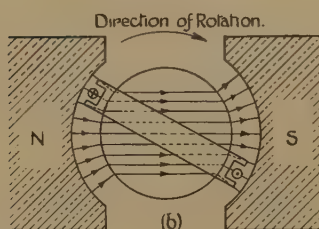
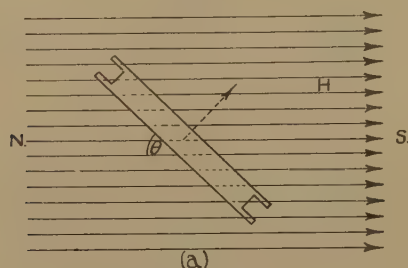
$$I\ddot{\theta} + \{\alpha + (A^2 N^2 H^2)/r\}\dot{\theta} + b\theta = 0,$$

or

$$\ddot{\theta} + \{\alpha + (A^2 N^2 H^2)/(rI)\}\dot{\theta} + \beta\theta = 0.$$

The effect of closing the galvanometer through a resistance  $r$  is therefore to increase the damping coefficient by an amount  $(A^2 N^2 H^2)/(rI)$ .

Fig. 24.



Measurement of high resistance by the method of damping.

Let  $\delta$  be the decrement per cycle (N.B.—not to be confused with  $\lambda$ , the decrement per half cycle). Then when  $r \rightarrow \infty$ ,  $\delta = \frac{1}{2}\alpha T$ , where  $T = 4\pi/\sqrt{4\beta - \alpha^2}$ , the period of the motion. When  $r$  has the values  $r_1$  and  $r_2$ , in turn, let the corresponding decrements be  $\delta_1$  and  $\delta_2$ , so that

$$\delta_1 = \frac{1}{2}\{\alpha + (A^2 N^2 H^2)/(r_1 I)\}T_1,$$

and

$$\delta_2 = \frac{1}{2}\{\alpha + (A^2 N^2 H^2)/(r_2 I)\}T_2,$$

where  $T_1$  and  $T_2$  are the new periodic times. In actual practice, on

account of the smallness of the total damping coefficient,  $T_1 \doteq T_2 \doteq T$ , so that we have

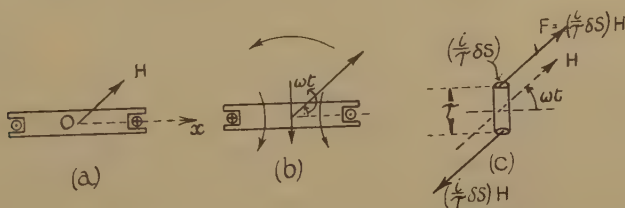
$$\frac{\delta - \delta_1}{\delta - \delta_2} = \frac{\lambda - \lambda_1}{\lambda - \lambda_2} = \frac{r_2}{r_1} = \frac{R_2}{R_1},$$

where  $R$  denotes the total resistance of the circuit in ohms. Hence it is only if  $G/R \rightarrow 0$ , where  $G$  is the resistance of the galvanometer in ohms, that this method is valid: of course,  $R$  must never be so low that the motion of the coil becomes non-periodic.

*On the Couple exerted on a Fixed Closed Coil when placed in a Rotating Magnetic Field.*

Let  $H$  be the constant magnitude of the magnetic field which rotates in an anti-clockwise direction with constant angular velocity  $\omega$ , time being measured from an instant when the field is directed from left to right and is parallel to the plane of the fixed coil. Then if a current flows in the coil its direction will be called positive if it is directed as indicated in fig. 25 (a). The flux of magnetic induction increases as the field rotates,

Fig. 25.



so that the field due to the induced current will have the direction indicated in fig. 25 (b). We assume that the coil consists of one turn in air, so that if  $A$  is its area, the flux,  $\Phi$ , through it at time  $t$  is  $AH \sin \omega t$ . The induced e.m.f. is therefore

$$e = -\frac{\partial \Phi}{\partial t} = -AH\omega \cos \omega t = AH\omega \cos (\omega t - \pi).$$

If  $l$  is the self-inductance of the coil and  $r$  its resistance (e.m.u.), the current,  $i$ , is given by

$$i = [(AH\omega)/(r^2 + \omega^2 l^2)^{1/2}] \cos (\omega t - \pi - \phi),$$

where  $\tan \phi = (\omega l)/r$ . Now this expression gives the magnitude and direction of the induced current: when the expression has a positive sign the current flows in the direction shown in fig. 25 (a). Hence the equivalent magnetic shell for each turn in the coil will have a pole strength per unit equal to  $(i/\tau)$ , where  $\tau$  is the thickness of the shell. The pole strength of an element of the shell of cross-sectional area  $\delta s$  is  $(i/\tau)\delta s$ . The couple on this element is

$$[(i/\tau)\delta s H \cos \omega t]\tau = iH \cos \omega t \delta s,$$

and this is the value of the couple in a clockwise direction. The couple in the direction of rotation of the magnetic field is therefore  $-iH \cos \omega t \cdot \delta s$ , and for the whole shell this becomes  $-AHi \cos \omega t$ .

Substituting for the value of  $i$  already obtained, we find that the instantaneous couple on the coil in the direction of rotation of the field is

$$AH\{AH\omega \cos (\omega t - \phi)/(r^2 + \omega^2 l^2)^{\frac{1}{2}}\} \cos \omega t.$$

The mean couple per cycle is therefore  $\frac{1}{2}\{(A^2 H^2 \omega)/(r^2 + \omega^2 l^2)^{\frac{1}{2}}\} \cos \phi$ , and this tends to cause the coil to rotate in the same direction as the field itself.

*Note on the Correction for Ferromagnetic Impurities in a Specimen whose Magnetic Susceptibility is being determined.*

Since very large magnetic fields are employed in determinations of the magnetic susceptibility of paramagnetic and diamagnetic substances, Honda and Owen \* assumed that any ferromagnetic impurity was magnetized to saturation. In such a case it seemed legitimate to assume that  $I$ , the "measured" intensity of magnetization, was given by

$$I = I_{\infty} + (I_{\text{Fe}})_{\text{sat.}},$$

where  $I_{\infty}$  is the intensity of magnetization for the pure specimen and  $(I_{\text{Fe}})_{\text{sat.}}$  is that of the ferromagnetic impurity. Now  $(I_{\text{Fe}})_{\text{sat.}}$  is constant, whereas  $I_{\infty}$  is a linear function of the magnetizing field,  $H$ . If  $\chi$  is the "measured" susceptibility,  $\chi_{\infty}$  that of the pure specimen, we have

$$\chi = \chi_{\infty} + \{(I_{\text{Fe}})_{\text{sat.}}\}/H.$$

Since  $\chi_{\infty}$  is a constant, if  $H^{-1}$  is plotted against  $\chi$ , the intercept of the straight line so obtained will give  $\chi_{\infty}$ , the quantity required.

Vogt † has shown that this last expression does not apply when the Gouy method is employed. The equation is only true for those methods in which the test object lies wholly in a field of such strength that the saturation of the ferromagnetic impurity is complete. In Gouy's method this is not so, for one end of the specimen lies in a magnetic field which is very small: let this be the upper end. If  $\alpha$  is the cross-sectional area of the specimen (in the form of a uniform rod), and  $\chi_a$  is the susceptibility of air, then, with the usual notation, the force  $\delta F$  acting upwards on a short length of the rod contained between planes at height  $z$  and  $z + \delta z$ , where the strength of the magnetic field is  $H$  and  $H + \delta H$ , respectively, is given by

$$\begin{aligned} \delta F &= \alpha \frac{\mu - \mu_a}{8\pi} \cdot 2H \cdot \delta H = \alpha(\chi - \chi_a)H\delta H \\ &= \alpha\chi H\delta H - \alpha\chi_a H\delta H \\ &= \alpha\chi_{\infty} H\delta H + \alpha\{(I_{\text{Fe}})/H\} H\delta H - \alpha\chi_a H\delta H, \end{aligned}$$

it no longer being assumed that  $(I_{\text{Fe}})_{\text{sat.}}$  is reached everywhere.

If  $mg$  is the weight of the load added to the balance pan to restore equilibrium,

$$mg = - \int_H^{H_0} dF = \alpha \frac{\chi_{\infty} - \chi_a}{2} (H^2 - H_0^2) + \alpha \int_{H_0}^H I_{\text{Fe}} dH.$$

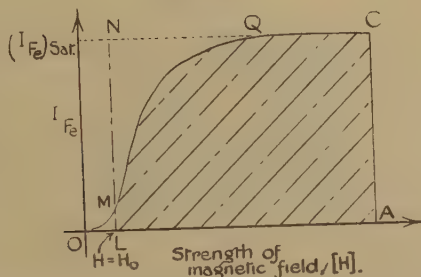
\* Honda, *Ann. der Phys.* xxxii. p. 1048 (1910); Owen, *ibid.* xxxvii. p. 657 (1912).

† Vogt, *Ann. der Phys.* xiv. p. 1 (1932).



To evaluate this last integral consider the magnetization curve for the impurity—cf. fig. 26. Then the curve OC shows how  $I_{Fe}$  varies

Fig. 26.



Magnetization curve for the iron impurity distributed throughout a specimen of non-ferromagnetic material.

with  $H$ . The shaded area represents the integral. Thus

$$\int_{H_0}^H I_{Fe} dH = \text{area LACQM} = (I_{Fe})_{sat.}(H - H_0) - \text{area MQN}.$$

But  $2mg/\{\alpha(H^2 - H_0^2)\}$  is  $(\chi - \chi_a)$ .

$$\therefore \chi - \chi_a = \chi_\infty - \chi_a + \frac{2(I_{Fe})_{sat.}}{H + H_0} - \frac{2 \text{ area MQN}}{H^2 - H_0^2}.$$

Since the last term tends to zero in all practical cases, we may write, if  $H_0 \rightarrow 0$ , where  $H_0$  is the field strength at the end of the specimen further from the strong field,

$$\chi = \chi_\infty + \{2(I_{Fe})_{sat.}\}/H.$$

Usually this final part of the proof is not mentioned. The above equation shows that if  $\chi$  is plotted against  $H^{-1}$ ,  $\chi_\infty$  is readily obtained.

### *On the Elementary Theory of the Kohlrausch Bridge.*

An elementary theory of the Kohlrausch bridge, shorn of all minor corrections, is as follows:—it would seem to be important since all the essential precautions which have to be observed when the bridge is in use are revealed. ABEF, fig. 27, is a diagrammatic representation of the bridge: the arm AB contains the electrolytic cell whose resistance is to be measured. Now the cell may be considered to be equivalent to a resistance  $R_1$  in series with a capacitance  $C_1$ , arising from the capacitances at the electrodes, the whole being shunted by a capacitance  $C_0$  to represent any stray capacitance in the cell. Then if  $\omega$  is the angular frequency of the source of alternating current and if  $Z$  denotes a vector impedance operator, we have, if  $j = \sqrt{-1}$ ,

$$\frac{1}{Z_1} = \frac{1}{-\frac{j}{\omega C_0}} + \frac{1}{R_1 - \frac{j}{\omega C_1}},$$

so that

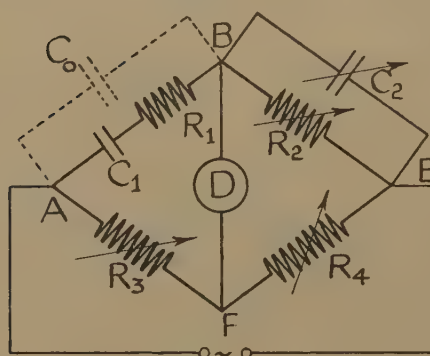
$$Z_1 = \frac{-\frac{j}{\omega C_0} \left[ R_1 - \frac{j}{\omega C_1} \right]}{R_1 - j \left[ \frac{1}{\omega C_1} + \frac{1}{\omega C_0} \right]}.$$

Also  $Z_2 = R_2 / (1 + j\omega C_2 R_2).$

Hence when the bridge is balanced

$$\frac{\left[ \frac{-\frac{j}{\omega C_0} \left( R_1 - \frac{j}{\omega C_1} \right)}{R_1 - \frac{j}{\omega C_1} - \frac{j}{\omega C_0}} \right]}{R_3} = \frac{R_2 / (1 + j\omega C_2 R_2)}{R_4}.$$

Fig. 27.



Equating real and imaginary parts

$$-\frac{1}{\omega^2 C_0 C_1} + R_1 R_2 \frac{C_2}{C_0} = R_1 R_2 \frac{R_3}{R_4},$$

so that it is only when  $\omega$  is large that  $C_2/C_0 = R_3/R_4$ ; also

$$\frac{-R_1}{\omega C_0} - \frac{C_2 R_2}{\omega C_0 C_1} = -\frac{R_2 R_3}{R_4 \omega C_1} - \frac{R_2 R_3}{R_4 \omega C_0}.$$

If  $\omega$  is large, so that  $C_2/C_0 = R_3/R_4$ , this equation yields

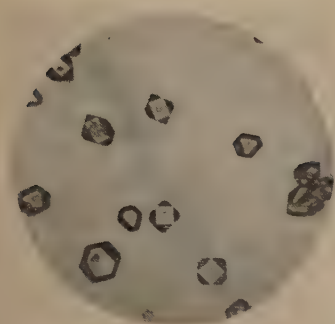
$$R_1 = R_2 R_3 / R_4,$$

which is the well-known relationship: but it is only when  $\omega$  is large that this equation is valid.

### On Mean Coefficients of Expansion.

*Note added February 27, 1946.*—If  $v$  is the volume of a fixed mass of substance at temperature  $t$ , we may write  $v=f(t)$ , and  $\alpha$ , the coefficient of expansion at temperature  $t$ , is, as usual, defined by the equation

$$\alpha = \frac{1}{v} \frac{dv}{dt}.$$



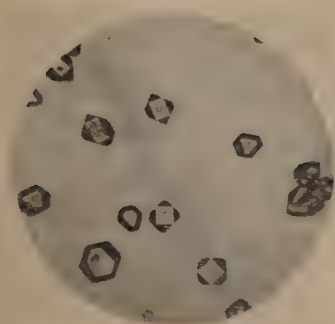
(a)



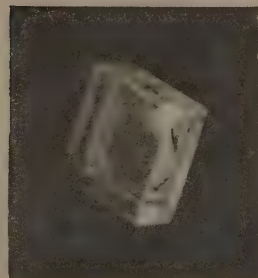
(b)



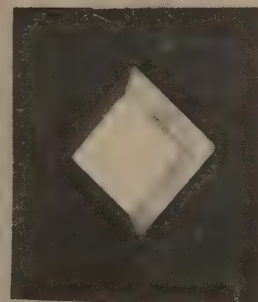
(c)



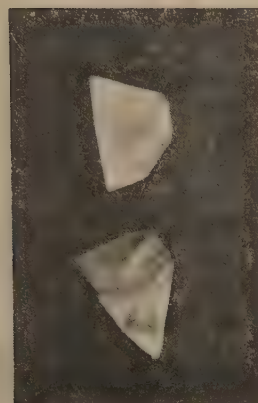
(d)



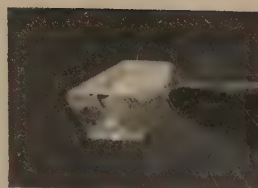
(e)



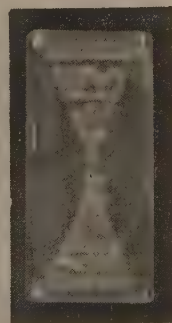
(f)



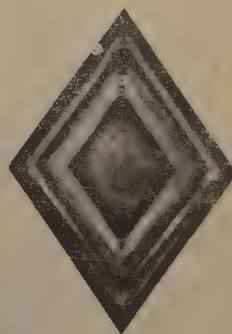
(g)



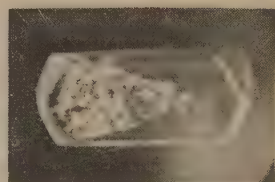
(h)



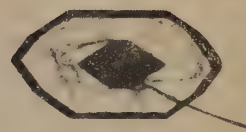
(i)



(j)



(k)



(l)

For description of Plate, see p. 629.





If  $\bar{\alpha}$  is the mean coefficient of expansion of the substance between temperatures  $t_1$  and  $t_2$ , ( $t_2 > t_1$ ), we have

$$\bar{\alpha} = \left( \int_{t_1}^{t_2} \alpha \, dt \right) / \left( \int_{t_1}^{t_2} dt \right) = \left( \int_{v_1}^{v_2} \frac{dv}{v} \right) / (t_2 - t_1) = \left( \log_e \frac{v_2}{v_1} \right) / (t_2 - t_1).$$

[Thus

$$\alpha = \frac{d}{dt} (\log_e v),$$

and although this may not be regarded as a satisfactory physical definition of  $\alpha$ , yet it is equivalent to it mathematically.]

The expression obtained for  $\alpha$ , the mean coefficient of expansion, is exact, but approximations to it are obtainable as follows. In the first place we may write

$$\log (v_2/v_1) = \log [1 + (v_2 - v_1)/v_1] = [(v_2 - v_1)/v_1] - \frac{1}{2}[(v_2 - v_1)/v_1]^2 + \dots,$$

so that if the *square* and higher powers of  $[(v_2 - v_1)/v_1]$  are neglected, we have

$$\bar{\alpha} = \frac{v_2 - v_1}{v_1(t_2 - t_1)},$$

which is generally accepted, but without proof.

A better approximation to the logarithmic formula for  $\bar{\alpha}$  is obtained as follows:

$$\log \frac{1+x}{1-x} = 2 \left[ x + \frac{x^3}{3} + \dots \right],$$

and hence if

$$\frac{v_2}{v_1} = \frac{1+x}{1-x}, \text{ or } x = \frac{v_2 - v_1}{v_2 + v_1},$$

we have, if the *cube* and higher powers of

$$\frac{v_2 - v_1}{v_2 + v_1},$$

are neglected,

$$\bar{\alpha} = 2 \cdot \frac{(v_2 - v_1)}{(v_1 + v_2)(t_2 - t_1)} = \frac{(v_2 - v_1)}{\frac{1}{2}(v_1 + v_2)(t_2 - t_1)}.$$

Similar expressions may be written down at once for the mean coefficient of linear expansion and also for the mean coefficient of increase of resistance with temperature, although this last coefficient is seldom, if ever, used.

## LXI. Crystal Habit and Internal Structure.—III.

By A. F. WELLS\*.

[Received July 29, 1946.]

[Plate XIV.]

### Introduction.

IN the present paper we retain as our working hypothesis the postulate of Gibbs that the equilibrium shape of a crystal is that of minimum total

\* Communicated by the Author.

surface free energy for a given volume, and we shall examine some of its implications in morphological crystallography. Although, owing to our ignorance of the absolute or even the relative values of the surface free energies of crystal faces, we cannot at present make the treatment quantitative, it is nevertheless possible to use this concept to clarify our ideas on the morphology of crystals. In Part I.\* we pointed out that in 20 of the 32 point-groups a general or special form encloses a polyhedron, whereas in the remaining point-groups faces of at least two types are required. We shall first consider what are the conditions for a crystal to be bounded exclusively by faces of one form. The Gibbsian treatment gives a precise answer to this question in terms of the relative surface free energies of the various possible faces, though, as we saw in Part I., the problem is complicated by the fact that for a complete solution all possible combinations of forms would have to be considered. To illustrate this point we shall consider the relative developments of  $\{100\}$  and  $\{111\}$  on a cubic crystal. Next, let us suppose that we know which form (or combination of forms) leads to the most stable shape for a crystal of a given substance. What is the significance of any other forms developed? This is a fundamental matter in crystallography because of the importance attached to the development of more complex forms when determining the point-symmetry of a crystal. Of particular interest are the "anomalous" developments of certain crystals, *i. e.* face developments which exhibit symmetry which is either greater or less than that of the internal structure. Examples of the former type are the symmetrical developments of crystals of polar classes, and of the latter the apparent point-symmetry 23 of lead and barium nitrates with the space-group  $P a3$ . It is clear that no complete correlation of crystal habit with internal structure will be possible until we gain some insight into these supposed anomalies. We shall discuss the following topics:—

1. The relative developments of different forms on crystals.
2. The Wulff theorem.
3. Face development and point-symmetry.
4. The effect of environment on crystal habit.
5. The effect of environment on apparent point-symmetry.
6. The genesis and significance of complex forms on crystals.
7. Some cases of anomalous face developments.

In later papers we propose to discuss in more detail the morphology of some of the substances mentioned in the present paper and to show how the method described in Section 2 leads for the first time to the experimental determination of the ratios of the surface free energies of certain crystal faces.

### 1. *The Relative Developments of Different Forms on Crystals.*

It might appear at first sight that if a face of one form is that of lowest surface free energy, and if that form encloses a polyhedron, then that

---

\* Previous papers in this series: I. *Phil. Mag.* xxxvii. p. 184; II. *Ibid.* xxxvii. p. 217.

form would appear exclusively on the crystal. This is not so. Since the geometrical relationships between the various sets of faces are different for the different crystal systems, we shall, for simplicity, deal with a cubic crystal, and investigate how the relative developments of cube and octahedron faces would depend on the relative surface free energies of these faces. The calculation would be relevant to a discussion of the habit of NaCl crystallizing from a solution containing NaOH or urea, or of alum from a solution containing certain dyestuffs.

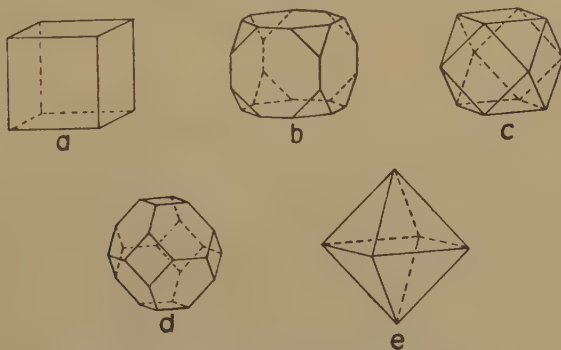
The change from cube to octahedron is illustrated in fig. 1. In (b) let the edge of the original cube be  $a$  and let the length  $y$  be removed by creating a (111) face on each corner. The volume of the solid figure (b) is

$$v = a^3 - 4y^3/3.$$

The total area of {100} is  $6(a^2 - 2y^2)$  and of {111},  $4\sqrt{3}y^2$ . If the surface free energies per unit area of these faces are  $g_{100}$  and  $g_{111}$ , the total surface free energy

$$G = 6(a^2 - 2y^2)g_{100} + 4\sqrt{3}y^2g_{111}.$$

Fig. 1.



The change from cube to octahedron.

For an equilibrated crystal the condition is that

$$\frac{\delta G}{\delta a} = \frac{\delta G}{\delta y} \quad \text{or} \quad \frac{12ag_{100}}{3a^2} = \frac{-24yg_{100} + 8\sqrt{3}yg_{111}}{-4y^2},$$

whence 
$$\frac{y}{a} = \frac{3g_{100} - \sqrt{3}g_{111}}{2g_{100}}.$$

For  $y=0$ ,  $g_{111} = \sqrt{3}g_{100}$ , i. e. if  $g_{111} > 1.73g_{100}$ , then the equilibrium shape is a cube. For  $y=a/2$ ,  $g_{111} = 2g_{100}/\sqrt{3}$ , the condition for the cubo-octahedron (c). For values of  $g_{111}/g_{100}$  between  $\sqrt{3}$  and  $2/\sqrt{3}$  the shape would be of type (b). To investigate the conditions for shapes between



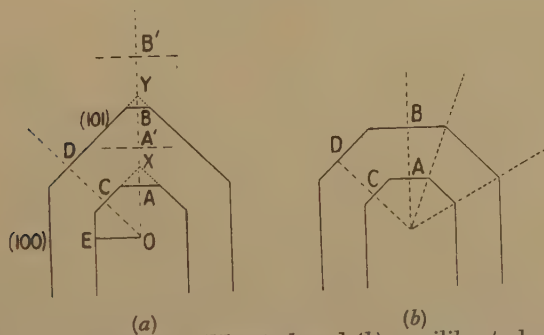


At various times measurements have been made of the relative rates of growth ( $v_{hkl}$ ) normal to different faces of a crystal. We can now examine in what circumstances such experiments can yield useful information about the relative surface free energies of the faces concerned. Let fig. 2 (a) represent, in elevation, two successive stages in the growth of an orthogonal crystal exhibiting  $\{100\}$ ,  $\{101\}$ , and  $\{001\}$ , and let the equilibrium shape be that completed by the dotted lines (to X and Y respectively), so that  $\{001\}$  is not an equilibrium form but has been produced on the crystal by some accidental cause. OCD is normal to  $(101)$ ,  $OB'$  to  $(001)$  and  $OE$  to  $(100)$ , O being the Wulff point. The broken lines parallel to  $(001)$  indicate the imaginary positions of  $(001)$  in the equilibrated crystal, *i. e.*  $OA' : OC : OE = g_{001} : g_{101} : g_{100}$ . We suppose now that the crystal grows from AC to BD and enquire what is the significance of, for example, the ratio  $AB : CD$ . Clearly,

$$\frac{AB}{CD} \neq \frac{A'B'}{CD} \equiv \frac{g_{001}}{g_{101}}.$$

(It might happen that  $AB = A'B'$  since both are greater than  $XY$ , but this would be fortuitous, for the position of A along OX has been determined by one or other of the influences discussed in Section 6.) In

Fig. 2.



Growth of (a) non-equilibrated and (b) equilibrated crystal.

fact,  $AB : CD = g_{001} : g_{101}$  only if the crystal grows as shown at (b), in which case the original crystal would have been an equilibrated crystal and  $\{001\}$  an equilibrium form. Thus the relative rate of deposition on an "unstable" face is greater than corresponds to its surface free energy, and accordingly such faces entirely disappear on further growth. We conclude that the relative rates of deposition on crystal faces can be used to determine ratios of surface free energies only if the faces are of equilibrium forms. We cannot determine  $g_{111}/g_{100}$  for NaCl from pure solution by producing artificially a  $(111)$  face and then comparing growth on this face with that on  $(100)$ . If, however, a crystal of NaCl is growing from a solution containing sufficient urea so that the equilibrium shape is bounded by  $\{100\}$  and  $\{111\}$ , then such relative growth measurements

will, if made on an equilibrated crystal, give the ratio  $g'_{111}/g'_{100}$  *relevant to that particular solution*. This ratio is not, however, the same as  $g_{111}/g_{100}$  for pure solution.

We can illustrate these points by reference to experiments of Spangenberg (1924) and France (1935). They can conveniently be considered together because the former measured  $v_{111}/v_{100}$  for NaCl crystals growing in solutions containing various amounts of urea, and the latter the ratio  $v_{100}/v_{111}$  for alum crystals growing from solutions containing various concentrations of the dyestuff diamine sky blue. For NaCl the equilibrium shape from pure solution is the cube and urea leads to the development of  $\{111\}$ , while in the case of alum the equilibrium shape from pure solution is the octahedron and the dyestuff is preferentially adsorbed on  $\{100\}$  leading to the development of cube faces. The ratios  $[v_{111}/v_{100}]_{\text{NaCl}}$  and  $[v_{100}/v_{111}]_{\text{alum}}$  should therefore show variations over similar ranges. Some experimental values are shown in Table I.

TABLE I.

Diamine sky blue g./100 cc. solution	$[v_{100}/v_{111}]$ alum		Urea, g./100 cc. solution	$[v_{111}/v_{100}]$ NaCl	
	Stirred	Unstirred			
0.000	1.75	1.61	0.0	3.19	$\pm 0.25$
0.002	1.66	—	3.5	2.20	0.30
0.004	0.75	0.38	8.0	0.68	0.03
0.006	0.79	0.29	13.1	0.32	0.02
0.008	0.45	0.00	21.9	0.32	0.03
			33.5	0.25	0.01

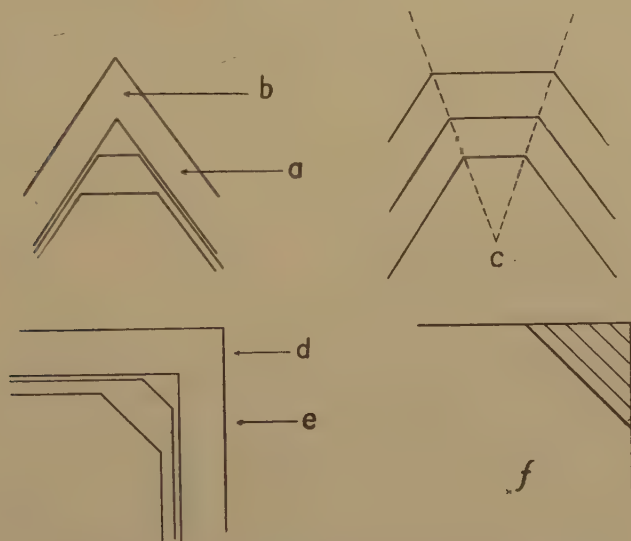
Values both within and outside the range 1.73—0.58 are observed. The types of growth to which they correspond are shown in fig. 3. A value of  $(v_{100}/v_{111})_{\text{alum}}$  greater than 1.73, fig. 3 (a), is observed if a crystal with cube faces grows in pure solution. We should expect this ratio to rise as growth continues, reaching the value  $\infty$  at the moment of completion of the octahedron. Thereafter the ratio has no meaning, since there is no (100) face, the crystal continuing to grow as an octahedron (b). If the concentration of dyestuff in the alum solution is such that the equilibrium shape consists of  $\{111\}$  and  $\{100\}$ , and the crystal on which measurements are made has the equilibrium shape, then growth takes place as at (c), when  $v_{100}/v_{111} = g_{100}/g_{111}$  and lies between 1.73 and 0.58. (If the original shape is different from the equilibrium shape some initial adjustment of shape is necessary, after which growth continues as at (c).) For all concentrations of dyestuff such that  $g_{100}/g_{111} < 0.58$  the crystal eventually grows as a cube (d). If a crystal with (111) faces is put into such a solution  $v_{100}/v_{111}$  is observed less than 0.58 (e), falling to the limiting

value zero, which is reached as the last trace of (111) disappears. The ratio  $v_{100}/v_{111}$  could also have the value zero over a finite time if the conditions were such that growth is completely inhibited on (100), *e. g.* by adsorption of dyestuff (*f*). As soon as the crystal is bounded entirely by {100} all growth ceases. The figures for the unstirred alum solutions containing dyestuff all lie outside the range 1.73–0.58 and therefore do not refer to growth of equilibrated crystals. The figure of 1.61 for the unstirred pure solution shows that here also “normal” growth (as an octahedron) was not taking place, *e. g.* growth may have been too rapid.

Our general conclusions are as follows:—

(i.) Observations of the growth of equilibrated crystals in any solution can lead directly to the relative surface free energies *appropriate to that solution* of faces present as equilibrium forms. If, as in the case of alum,

Fig. 3.



Possible types of growth of alum crystals (see text).

sodium chloride and chlorate, barium nitrate, etc., the equilibrium shape from *pure* solution consists of only one form, then no information about surface free energies can be obtained directly from such measurements.

(ii.) The surface free energies of other faces (not present on the equilibrated crystal from pure solution) relative to that of the equilibrium form might be obtained in favourable cases as follows: If by the addition of a suitable foreign material the required face can be developed, then measurements are made of, say,  $(v_{111}/v_{100})_{\text{NaCl}}$  for various concentrations of urea, within the range of urea concentrations for which the equilibrium shape consists of both {100} and {111}. (Spangenberg states that from the highest urea concentrations down to 8.5 g./100 cc. solution the final crystal was an octahedron; for 8.5–5.0 g. it consisted of {100}+{111}, and below 5.0 g. urea a simple cube resulted. For this system the

required range of urea concentrations is therefore 8.5–5.0 g./100 cc. solution.) These measurements must be made in each case for the growth of an equilibrated crystal, when  $v_{111}/v_{100}=g_{111}/g_{100}$ , *i. e.* the value of the ratio is significant only when it maintains a constant value characteristic of a particular concentration of urea. Then by extrapolation it should be possible to obtain  $v_{111}/v_{100}$  for zero concentration of urea. This gives a general method for determining the relative surface free energies of faces other than those present on the equilibrated crystal from pure solution, provided that the face(s) can be made to appear by some alteration in the composition of the solution.

### 3. Face Development and Point-symmetry.

Crystallographers have always been interested in the development of faces with high indices because of their diagnostic value. In a particular point-group the operation of the symmetry elements leads to a set of related faces (a form), and if the face is of the most general form  $\{hkl\}$  the number of, and angular relations between, these related faces are characteristic of the point-group. In the cubic point-groups the general forms are as follows :

				No. of Faces.
23	tetrahedral	pentagonal	dodecahedron	.. 12
43	pentagonal	icositetrahedron	.. ..	24
<i>m</i> 3	dyakis	dodecahedron	.. ..	24
43 <i>m</i>	hexakis	tetrahedron	.. ..	24
<i>m</i> 3 <i>m</i>	hexakis	octahedron	.. ..	48

These general forms differ from one another in number and/or arrangement of faces. If  $\{hkl\}$  consists of 24 faces, the point-group is 43, *m*3 or 43*m*, and the arrangement of the faces would distinguish between these three possibilities. In other cases a knowledge simply of the *number* of equivalent faces ( $hkl$ ) should distinguish certain point-groups from others. For example, simply as regards face development, the only difference between the point-groups 43 and *m*3*m* is that in the former  $\{hkl\}$  is a pentagonal icositetrahedron and in the latter a hexakis octahedron. The latter figure could occur in 43 only if two sets of (non-equivalent) faces  $\{hkl\}$  and  $\{h\bar{k}l\}$  were developed to the same extent. Alternatively a pentagonal icositetrahedral development could result in *m*3*m* if one-half the  $\{hkl\}$  faces developed to a different extent from the others—in the limiting case, one-half being absent. Since there would appear to be no reason for such unsymmetrical development of equivalent faces or such symmetrical development of non-equivalent faces, it has been generally assumed that the face development of a crystal will indicate the true point-symmetry. Now, since  $\{hkl\}$  will not usually have developed during normal growth of the crystal (see Section 6), *i. e.* the crystal is not an equilibrated crystal, it is worth while examining this matter in more detail. The point-group 23 is converted into *m*3 by the addition of planes of symmetry parallel to (100), (010) and (001). If,



therefore, the conditions of growth of the crystal were such that the symmetrical development of faces related by these symmetry planes was not possible, it might permit the crystal with true point symmetry  $m\bar{3}$  to exhibit only 23. The point-groups  $4\bar{3}$  and  $m\bar{3}m$  are related in a similar way. It might then be significant that the following examples of "anomalous" face developments are of this type:—

	Point-symmetry deduced from		Space-group
	Morphology	Internal structure	
Lead nitrate Barium nitrate etc. }	23	$m\bar{3}$	P $a\bar{3}$
Cuprous oxide Ammonium chloride Potassium chloride }	43	$m\bar{3}m$	<div style="display: inline-block; vertical-align: middle;"> <math>\left\{ \begin{array}{l} \text{P } n\bar{3}m \\ \text{P } m\bar{3}m \\ \text{F } m\bar{3}m \end{array} \right.</math> </div>

(Miers ('Mineralogy,' 2nd edition, p. 55) states that artificial pyrites sometimes shows tetrahedral habit. The space-group is P  $a\bar{3}$ .)

We shall return to this point in section 5.

So far we have considered only general forms  $\{hkl\}$ . Faces of this sort do not, however, always occur. In fact, as we shall see later, there is good reason to think that they would rarely be observed on equilibrated crystals. In practice, therefore, point-symmetry is not always deduced from the number and arrangement of faces of complex forms. It becomes necessary to take account of physical properties (optical activity, pyro-electricity, etc.), the orientation of etch-pits on different faces, and of the *relative areas* of faces. For example, if four (111) faces, forming a positive tetrahedron, are consistently developed to a greater extent than the other four (negative tetrahedron), this is regarded as evidence of tetrahedral rather than octahedral point-symmetry. Now, as soon as we begin to accept as evidence for point-symmetry differences between areas of faces, we are on less certain ground. In many cases there is, of course, no doubt that equivalent faces have become unequally developed simply as an accident of growth. For example, a "flattened" cube, point symmetry  $4/m\bar{m}m$ , can be safely dismissed as a distorted cube because it is isotropic. Since crystals are so commonly grown resting on one face, we shall first consider how this affects their shape, and then show that in certain cases the unsymmetrical environment can lead to crystals whose apparent point-symmetry is lower than the true point-symmetry.

#### 4. The Effect of Environment on Crystal Habit.

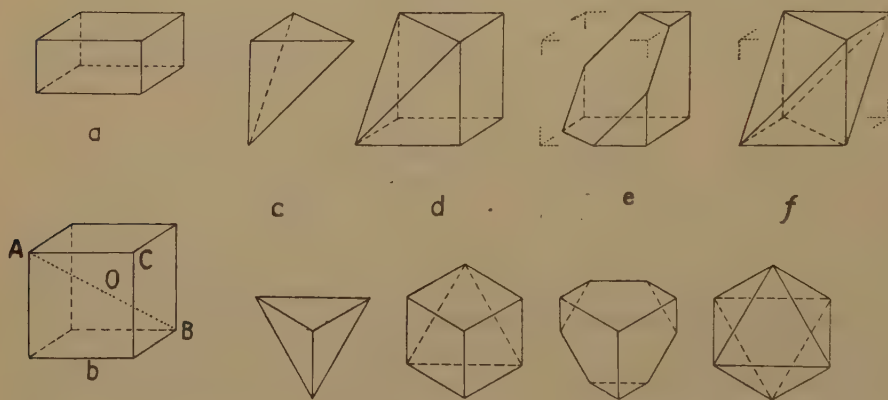
Neglecting any other complicating factors, the rate of growth on a crystal face will be greater the greater the surface free energy per unit area. If we prevent access of the solution to a given face, growth on that

face is prevented, *i. e.* as regards the growth of the crystal its effective surface free energy is zero. If therefore one face of a crystal is resting on the bottom of a dish or is in the surface of the solution, there is the possibility that the shape of the crystal may be different from that of a crystal surrounded on all sides by solution. The reduction of the surface free energy of one face to zero implies that the Wulff point lies in that face, so that the resultant shape may easily be derived. The condition then to be satisfied is that equivalent faces are equidistant from the Wulff point. We shall consider three simple figures, the cube, octahedron and tetrahedron, resting on (100) or (111) faces.

*Cube.—Resting on (100):* The expected form is the half-cube, fig. 4 (a).

*Resting on (111):* In fig. 4 (b) points between A and O are equidistant from three (100) faces, and points along OB are equidistant from the other three (100) faces. Only the point O is equidistant from all six cube faces. The "half-cube" (e) has the same surface free energy : volume

Fig. 4.



Shapes of crystals resting on one face—cube. The shapes (c)–(f) are also shown in plan, resting on (111).

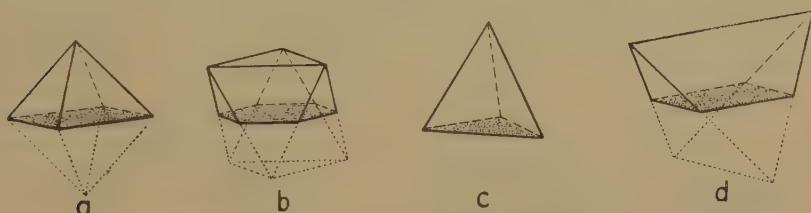
ratio as the cube itself, counting  $g_{111}=0$ . The shapes obtained by moving the (111) plane are of some interest. If the corner of (111) lies between A and C in fig. 4 (b), the condition that equivalent planes are equidistant from the Wulff point is satisfied but the crystal has a greater surface free energy : volume ratio than (e). Thus (c) has area : volume =  $9/a$  (where  $a$  is the cube edge), as compared with  $6/a$  for the cube. The complementary portion of the cube (d) has a smaller area : volume ratio than the cube (in this case  $5.4/a$ ) but is not a stable form because all six cube faces are not equidistant from the Wulff point, which must lie in the (111) face. Actually, shapes (c) and (d), in addition, of course, to (e), have been observed when solutions of NaCl and  $\text{NaClO}_3$  crystallize on a microscope slide, or when crystals grow downwards from the upper surface of their solutions. Occasionally also the (unstable) shape (f) occurs. (It is worth noting that when viewed along a three-fold axis, as when they are growing

on (111) on a slide, the shapes (c) and (f) are indistinguishable from regular tetrahedra and octahedra respectively. It is necessary to view them along [100] to see that certain of the inter-edge angles are  $90^\circ$  and not  $60^\circ$ .)

*Octahedron.*—The forms here are the “half-octahedra” shown in fig. 5, (a) and (b), according as the crystal is resting on (100) or (111) respectively. Crystals approximating to both these shapes are commonly observed in the case of alum, barium nitrate, etc.

*Tetrahedron.*—In a tetrahedron suspended in solution the Wulff point is at the geometrical centre. For a tetrahedron resting on one of its faces, fig. 5 (c), the Wulff point is now in the base and the total surface free energy of the crystal is only three-quarters of that of a tetrahedron surrounded on all sides by solution. For a crystal resting on a cube face the “half-tetrahedron” (d) has the same surface free energy: volume ratio as the normal tetrahedron. If  $\text{NaClO}_3$  is crystallized from a solution containing  $\text{Na}_2\text{S}_2\text{O}_3$ , crystals of both types (c) and (d) are obtained (see Pl. XIV. and Section 7 (a)).

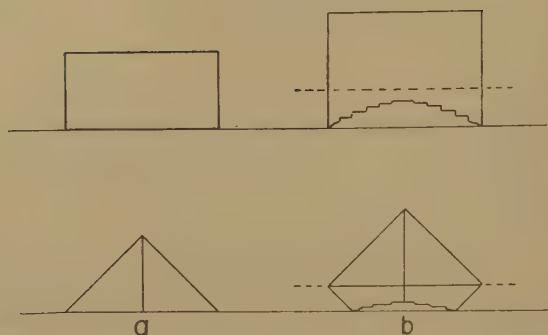
Fig. 5.



Shapes of crystals resting on one face—octahedron and tetrahedron.

Similar considerations could be applied to crystals of any other shape, the crystal being cut by a plane passing through the Wulff point, and many such shapes have been observed in the course of crystallization experiments. The face in the glass-solution or air-solution interface may be one which, if in contact with the solution, would have a very high surface free energy. Thus we find crystals of pentaerythritol growing downwards from the surface of a supersaturated solution with (001) in the surface, anthranilic acid growing on (010), and so on, these faces being highly unstable in contact with the solution. Crystals growing on the bottoms of vessels often have shapes approximating to those described above if contact between the undersurface of the crystal and the glass is good. The adhesion of the crystal to the glass depends not only on the planeness and cleanness of the glass but also on the nature of the crystal, and it is common experience to find that considerable force is required to remove a crystal. In other cases, the crystal is only lightly resting on the glass surface, and although growth on the lower surface is hindered it does take place slowly. We then observe a characteristic concave or terraced face and a shape intermediate between the “half-crystal” and the normal crystal (fig. 6).

Fig. 6.

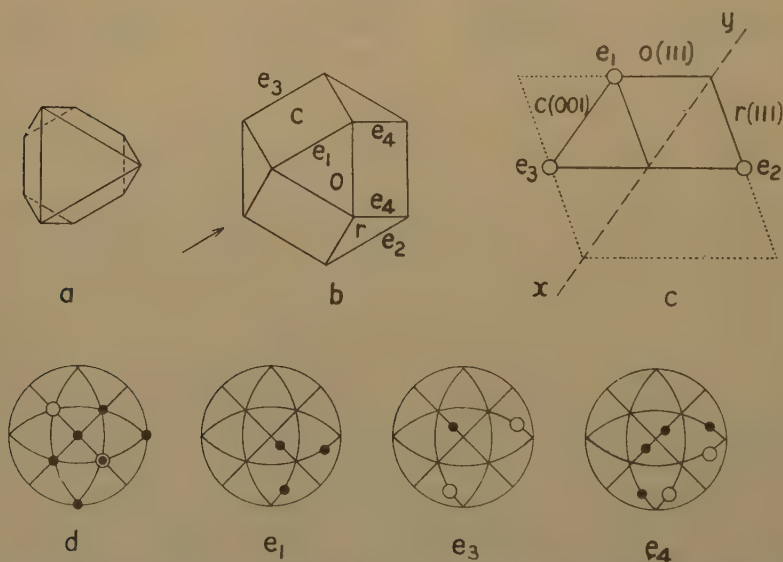


Shapes of crystals resting on one face: (a) good contact between crystal and glass surface, (b) poor contact.

### 5. The Effect of Environment on Apparent Point-Symmetry.

Consider the growth of a half-octahedron of a substance such as barium or lead nitrate or potassium alum (class  $m3$ ) resting on a (111) face. Over a period of days (100) faces appear, fig. 7 (b), and later, in the case of lead nitrate, pyritohedron faces. The planes of symmetry are parallel to the

Fig. 7.



The growth of a crystal resting on (111). (c) is the elevation of the crystal shown in plan at (b), and (d) its stereogram. The stereograms  $e_1$ ,  $e_3$  and  $e_4$  indicate the poles of faces truncating the corresponding edges in (b).

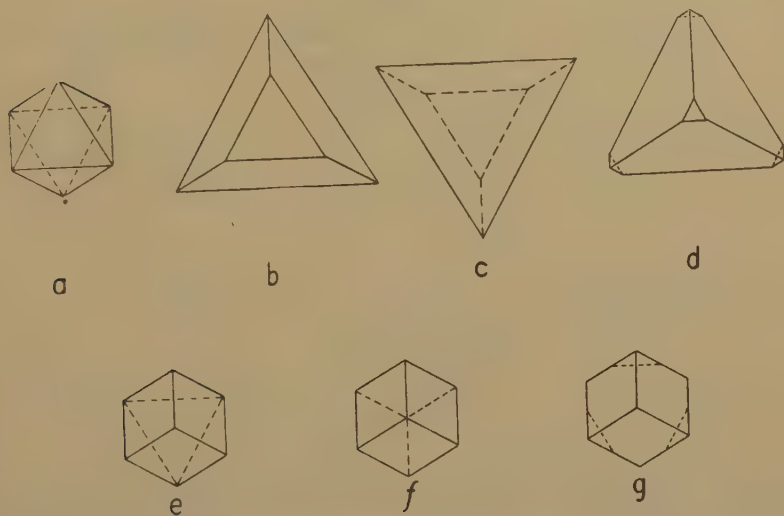
cube faces. At (c) is shown an elevation of crystal (b) viewed in the direction of the arrow, the plane of symmetry parallel to the face  $c$  being  $XY$ . This plane of symmetry relates  $o$  to  $r$ , and the edge  $e_1$  to  $e_2$ . On the half-octahedron these edges are clearly no longer equivalent. Instead of there being 24 equivalent edges between (111) and (100) faces as on a



normal truncated octahedron, there are only 12 on the crystal (*b*) and these fall into two sets of 3 each and one set of 6. The truncation of these edges will give rise to sets of faces as shown in the stereograms  $e_1$ ,  $e_3$  and  $e_4$ , which may well develop to different extents, so that the effect of growing on a (111) face is the apparent loss of the three planes of symmetry. The point-symmetry would thus appear to be 23 rather than  $m\bar{3}$ , or 43 instead of  $m\bar{3}m$ . The fact that the cases of anomalous cubic crystals quoted above are precisely of this sort does suggest that such effects may occur.

It is interesting also to consider the further growth preferentially along one of the 3-fold axes, giving crystals of trigonal symmetry. These, like most crystals described in the literature, are not equilibrated crystals, but they do occur, and it has not always been appreciated that deductions about point-symmetry should not be drawn from their shapes. Some

Fig.



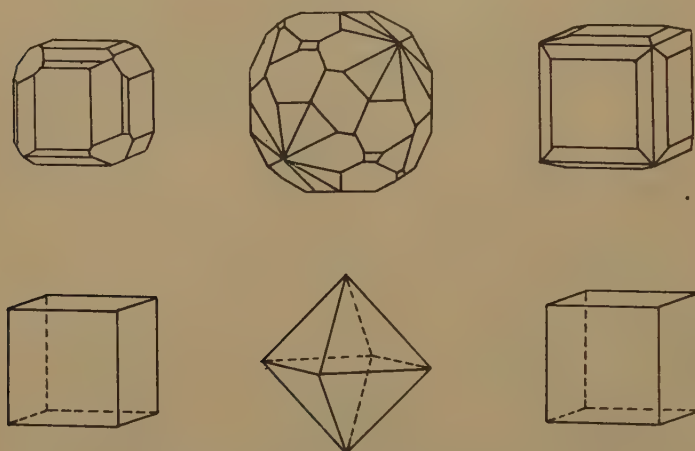
Growth of an octahedron resting on (111) to shapes of lower symmetry

simple shapes resulting from different rates of growth on various sets of faces of an octahedron are shown in fig. 8. They correspond to preferential growth on : (*b*) the three underneath faces, (*c*) the three upper faces, (*d*) the three underneath faces and the top face, (*e*) the top face, and (*f*) the top and bottom faces. Crystals of a markedly tetrahedral development as (*b*)–(*d*) are commonly observed when barium or lead nitrate crystallizes (rapidly) on a microscope slide. The rhombohedron (*f*) and the “half-rhombohedral” (*g*) are interesting, for crystals of barium nitrate of shape (*g*) and shapes intermediate between (*f*) and (*g*) can be grown (Pl. XIV. (*h*)), particularly when growing downwards from the surface. A simple calculation shows that (*f*) (or (*g*) if the surface free energy of the hexagonal base is counted as zero) has a total surface free energy only 7 per cent. greater than an octahedron with the same volume.

### 6. *The Genesis and Significance of Complex Forms on Crystals.*

A striking feature of the descriptions of crystals, both natural and artificial, is the complexity of face development. This gives an entirely misleading picture of what we may call the fundamental morphology of the crystals. Compare, for example, the illustrations in fig. 9 taken from the literature with the equilibrium shapes of the crystals grown slowly. In fact, the remarkable feature of the crystallography of many substances is not the complexity but the simplicity of their face developments, if the crystals are grown carefully to avoid the complications discussed in this section. The facts that NaCl and  $\text{NaClO}_3$  grow as simple cubes,  $\text{NaNO}_3$  as rhombohedra, and the alums and lead nitrate as octahedra, are very significant in throwing light on the relative values of the surface free energies of various possible faces. Thus the cubic development of NaCl means that  $g_{110} > \sqrt{2}g_{100}$ ,  $g_{111} > \sqrt{3}g_{100}$ , etc., *i. e.* there would appear to

Fig. 9.



Above : Crystals of  $\text{NaClO}_3$ ,  $\text{Ba}(\text{NO}_3)_2$  and NaCl from Groth.

Below : Equilibrated crystals of the same salts from aqueous solutions.

be very considerable differences between the surface free energies even of these simple forms. This is important in connection with the development of still more complex forms, particularly certain cubic forms such as the four-faced cube (fig. 10 (a)). In fig. 10 (b), which is the plan of a four-faced cube  $\{310\}$ , OA and OCB are normal to  $(310)$  and  $(100)$  respectively, where  $\text{OC} : \text{OA} : \text{OB} = 1.00 : 1.27 : 1.33$ . From these values we derive the following conditions for the development of  $\{100\}$  and/or  $\{310\}$ :

$$g_{310}/g_{100} \longrightarrow 1.27 \longleftarrow 0.95 \longleftarrow$$

$$\{100\} \text{ only} \qquad \{310\} + \{100\} \qquad \{310\} \text{ only,}$$

with which may be compared

$$g_{111}/g_{100} \longrightarrow 1.73 \longleftarrow 0.58 \longleftarrow$$

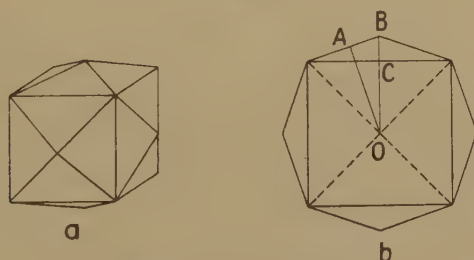
$$\{100\} \text{ only} \qquad \{100\} + \{111\} \qquad \{111\} \text{ only.}$$

Thus  $\{310\}$  can appear exclusively only if  $g_{310} < 0.95g_{100}$ , and as a form bevelling a cube only if  $1.27g_{100} > g_{310} > 0.95g_{100}$ . Since, in general, we should expect  $g_{310} \gg g_{100}$ , it is clear that the appearance of  $\{310\}$  as an *equilibrium form* is highly improbable. (It appears not infrequently on natural crystals of fluorspar.)

We shall accordingly postulate that the equilibrium shape of a crystal is likely to be a relatively simple development. This has been found to be true in those cases (at present relatively few in number) where the shape of the equilibrated crystal is known (*e. g.* those in the lower part of fig. 9). Equilibrated crystals may be of the following types :—

(i.) All the faces are of one form. This is possible in 20 of the 32 point-groups (see Part I.), in which the general or special form encloses a polyhedron. The indices of the faces may be simple (*e. g.* 100 in the cubic or rhombohedral systems) or complex (*e. g.* a rhombic bipyramid  $\{hkl\}$ ). It should be noted that the fact that a crystal is bounded by only one form does not necessarily mean that this is the equilibrium form (*cf.* the four-faced cube, above).

Fig. 10.



The four-faced cube.

(ii.) Two or more forms are present. In some point-groups at least two forms are necessary to form a closed figure, but even in one of the above-mentioned 20 point-groups two or more *special* forms may be required for the same reason (*e. g.*  $\{110\}$  and  $\{001\}$  of a holohedral orthorhombic crystal, this combination being more stable than, say, the bipyramid  $\{111\}$ ). In other cases two or more forms are present because, although each separately would enclose a polyhedron, together they form a solid with lower total surface free energy for a given volume than either form separately, as in the case of a cubic crystal if  $g_{111}/g_{100}$  lies between 1.73 and 0.58.

When other forms are present in addition to those of the equilibrated crystal they are the result of one or more of a number of disturbing influences to which the crystal has been subject during growth. This simplifies the general picture because we shall not have to account for the extraordinary variety of face developments in terms of surface free energies. On the other hand, it means that we shall have to suggest other reasons for the appearance of these more complex forms.

*The Effect of Impurities.*

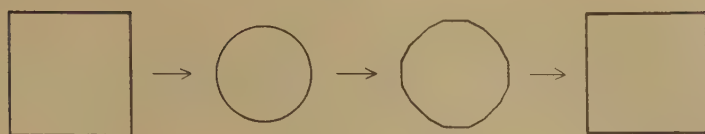
The presence of impurities in the solution (or melt) is probably a very important reason for the complex face developments of mineral crystals. Since suitable substances added to a solution can change  $g_{100}$  for a cubic crystal from a value greater than  $1.73g_{111}$  to a value less than  $0.58g_{111}$ , it is reasonable to suppose that at intermediate concentrations the same impurity can lead to values between these extremes, when both  $\{100\}$  and  $\{111\}$  will occur. Now, provided the concentration of impurity is maintained constant, the shape of the crystal may be regarded as the equilibrium shape *for that particular solution* (or melt), and not as abnormal. It must be made clear that a crystal showing faces not present on an equilibrated crystal *from pure solution* is not necessarily non-equilibrated. If the crystal has maintained the same shape throughout growth it is an equilibrated crystal. However, if the composition of the solution or melt has changed appreciably during the growth of the crystal, as has probably been the case with many minerals, then the crystal is likely to show a complex development which we cannot expect to explain in terms of the relative surface free energies of the face present. The complex development represents an intermediate phase in the growth of a crystal which has not had time to reach equilibrium with its surroundings. Finally, even if the environment of a mineral crystal had finally remained constant long enough to allow equilibrium to be reached, we cannot hope to account for its shape because we do not know the composition of the solution or melt with which it was finally in equilibrium.

*The Effect of Partial Dissolution during Growth.*

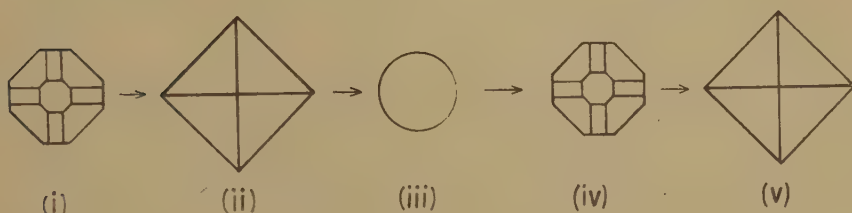
The initial localized dissolution of a crystal results in the formation of etch-pits on the faces. These may be regarded as "inverse vicinal faces," often belonging to forms with very high indices. Further dissolution leads to rounding-off of edges and corners. Regrowth does not in general lead directly to regeneration of the original faces but first to the development of faces truncating or bevelling the original edge or corner. Thus, if a cube of NaCl is dissolved away to a sphere, it grows as shown in fig. 11 (a), as demonstrated by the beautiful experiments of Neuhaus (1928). Similar results may be obtained with alum. The crystal (i) in fig. 11 (b) with  $\{111\}$ ,  $\{110\}$  and  $\{100\}$  was from a batch grown by leaving a solution to evaporate on the bench. Further slow growth, without temperature fluctuations, led to the octahedron (ii). This was then partially dissolved to the rounded object (iii), which on further slow growth gave (iv), with similar face development to (i), and later the octahedron (v). This clearly shows that the relatively complex development of (i) was due to the temperature fluctuations during its growth which resulted in partial dissolution and regrowth. The development shown in fig. 7 (b) arose in the same way. Repetition of this alternate dissolution and regrowth can obviously lead to crystals with the complex shapes so often found in the literature.



Fig. 11.



□



b

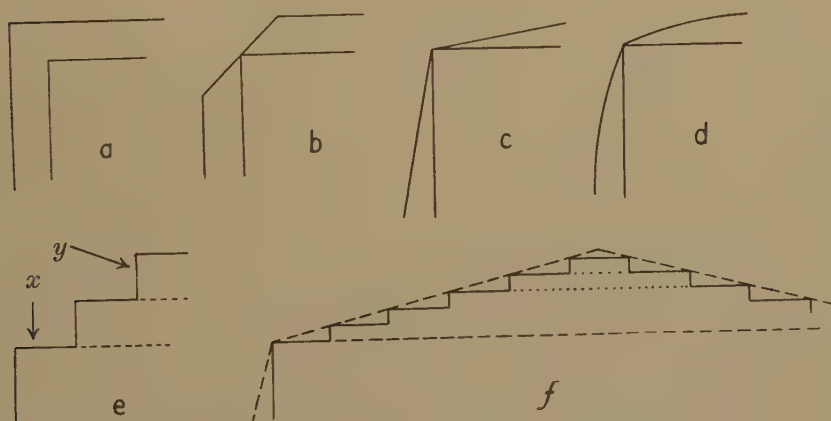
(a) Partial dissolution and regrowth of a cube of NaCl.

(b) Alternate growth and dissolution of a crystal of alum (see text).

*Non-completion of Layers on Certain Faces.*

A second type of growth leading to the development of complex forms is as follows:—Consider the deposition of new material on two adjacent

Fig. 12.



Possible types of growth on faces of a crystal.

faces of, for simplicity, a cube. Normal growth of the crystal is shown in fig. 12 (a). Other possibilities are shown at (b), (c) and (d). In each case deposition is assumed to take place only on the cube faces. The

results depend on the way in which successive layers terminate, leading to :

(b) Truncation of the original edge by (110).

(c) Beveling of the corner. The faces may either have simple indices, *e. g.* the four-faced cube {310}, or they may be vicinal faces.

(d) Curved faces.

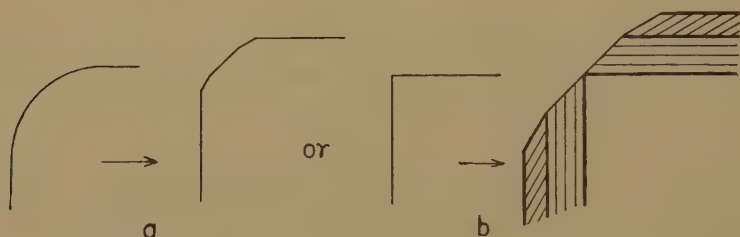
From fig. 12 (e) the conditions are seen to be :

$y/x$  constant  $\begin{cases} y/x \text{ is a simple fraction, } e. g. 1/1 \text{ for } (110), 1/3 \text{ for } (310). \\ y/x \text{ is very small, resulting in vicinal faces or} \\ \quad \text{"tapered" crystals.} \end{cases}$

$y/x$  decreasing  $\rightarrow$  curved face.

The growths observed in practice may be simply (b), (c) or (d), but more often represent combinations of these with (a), *i. e.* with some overall growth of the crystal. For example, vicinal faces are always present on an octahedron of alum which, however, still grows larger with essentially octahedral shape. The vicinal faces are observed at any given stage in the growth because before a particular layer has been completed to the edges of the face subsequent layers have already started from the centre

Fig. 13.



Two possible ways in which complex faces can arise.

of the face, and we have the state of affairs shown (in exaggerated form) in fig. 12 (f). Miers (1904) thought it paradoxical that the vicinal faces, which are faces of such low reticular density, should occur in preference to the much more densely populated (111) faces. The "paradox" is resolved when we realize that growth does not take place *on* the vicinal faces at all, but that they arise merely incidentally as the result of the particular mechanism of deposition of material *on* (111). In this sense the vicinal face (251, 251, 250) is not a face but the contour of a series of steps parallel to (111).

Thus, whether we regard a complex shape as having developed as shown in fig. 13 (a) or as in (b), it is clear that in (a) further normal growth will lead to greater deposition on the complex face than on the simple face, leading ultimately to the disappearance of the former, and in (b), deposition has never taken place on the complex face at all, and if it did, the complex face would disappear as in (a). We cannot discuss in detail the reasons for growth of the types shown in fig. 12. They fall

outside the scope of our present discussion of the growth of equilibrated crystals. The following examples will suggest some of the possible reasons for these abnormal types of growth: Miers (*loc. cit.*) showed that the angles between vicinal faces varied with the depth to which the crystal was immersed in a solution containing excess solid at the bottom; crystals of pentaerythritol and  $\text{KH}_2\text{PO}_4$  tapering to a point may be grown from highly supersaturated solutions, and in the case of  $\text{NH}_4\text{H}_2\text{PO}_4$ , tapered crystals with plane faces may be grown from solutions containing small concentrations of foreign metal ions (Kolb and Comer, 1945); curved faces are often produced in the presence of habit-modifying impurities, *e.g.* cubes of alum with curved faces from solutions containing bismarck-brown, and tetrahedra of sodium chlorate with curved faces from solutions containing sodium thiosulphate.

We are now in a position to reconsider briefly the "experimental morphological aspect" (E.M.A.) of Donnay and Harker derived in the manner described by Tremblay (1942). (1) The E.M.A. derived from observations on minerals is not necessarily related in any way to the relative stabilities of faces on crystals grown from pure solution. Each solution contained different amounts of different impurities so that it is not permissible to combine together observations made on crystals from different localities. (2) The most favourable case would be to suppose that we had numerous crystals of, *e.g.* NaCl with different developments all grown from pure solution. The equilibrium form for crystals from pure solution is {100}. The development of any other forms is due to one or more of the influences discussed above, so that a crystal other than a cube would not be stable in pure solution but on further growth would eventually become a simple cube. We might perhaps expect a very rough (inverse) relationship between the relative areas and the relative surface free energies of the more complex forms on the grounds that the less stable the face the less chance there is of observing it, because the greater its surface free energy the more rapidly it will disappear if growth of the crystal takes place. However, the greater the extent to which edges and corners are truncated or bevelled by complex faces the smaller becomes the area of {100} left, so that this rough relationship between the areas of faces and their surface free energies could only hold if the degree of truncation or bevelling is, on the average, small, *i.e.* so that {100} will always be the predominant form. (3) In comparing the "experimental" with the "theoretical" M.A. we are faced with the additional complication that the latter in many cases does not agree with the observed face developments in cases where there is no doubt about the equilibrium shape of the crystal. Any detailed agreement between these experimental and theoretical morphological aspects is therefore fortuitous.

### 7. Some Cases of Anomalous Face Developments.

Corresponding to the 32 point-groups there are 230 space-groups, each point-group giving rise to one or more of the latter. The space-groups

are more numerous than the point-groups because of the possibility (1) of combining the same symmetry elements with different lattice types, and (2) of the existence of symmetry elements involving translations (screw-axes and glide-planes), symmetry elements which have no meaning when applied to the face development of a crystal. Thus the point-group  $2/m$  gives rise to six space-groups,  $P 2/m$ ,  $C 2/m$ ,  $P 2_1/m$ ,  $P 2/c$ ,  $P 2_1/c$  and  $C 2/c$ . It is assumed that a crystal with any one of these space-groups will exhibit the point-symmetry  $2/m$ . This assumption that the morphological symmetry of a crystal should correspond to the point-group from which its space-group is derived implies that, owing to the regular repetition of the unit cell, screw-axes and glide-planes become equivalent to simple rotation axes and mirror planes as regards the macroscopic symmetry of the crystal. The space-groups derived from any point-group except 1,  $\bar{1}$ ,  $\bar{3}$ ,  $\bar{4}$ , and  $\bar{6}$  fall into two groups: (i.) those in which the point-symmetry of each lattice point is that of the point-group, and (ii.) those in which it is lower, owing to the presence of screw-axes or glide-planes. For example:—

Point-group.	Space-groups.	
	Type (i.).	Type (ii.).
$2/m$ $m\bar{3}$	$P 2/m$ , $C 2/m$ $P m\bar{3}$ , $F m\bar{3}$ , $I m\bar{3}$	$P 2_1/m$ , $P 2/c$ , $P 2_1/c$ , $C 2/c$ $P n\bar{3}$ , $F d\bar{3}$ , $P a\bar{3}$ , $I a\bar{3}$

The total of space-groups of the first type is 73. The question arises: Should a crystal with the space-group  $P a\bar{3}$ , for example, show the same face development as one with space-group  $P m\bar{3}$ ? A regular octahedron of  $Pb(NO_3)_2$  differs in a rather subtle way from an octahedron of a substance with space-group  $P m\bar{3}$ . Fig. 14 shows the atomic arrangement in the surfaces of the former, and it will be seen that the surface structures of adjacent octahedron faces are not symmetrical across the edges of the octahedron. However, any isolated area of one octahedron face is identical with that of any other, *i. e.* the surface free energies per unit area of all the (111) faces are equal, and there is no evidence that this lack of symmetry has any effect on the growth of the crystals.

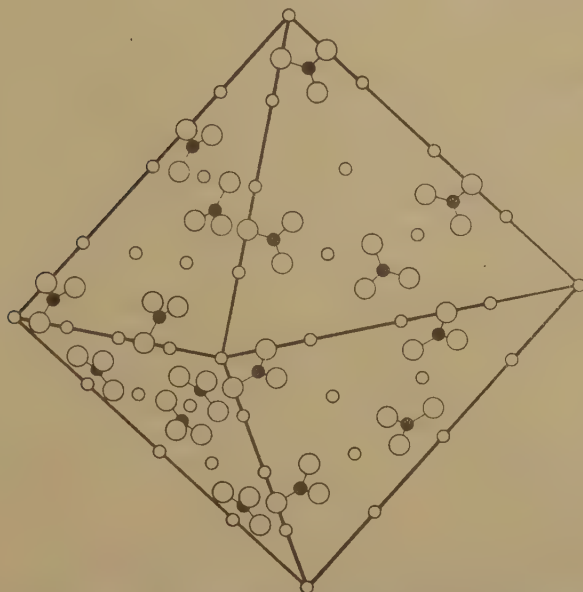
In this section we shall discuss (a) certain cubic crystals, the face developments of which apparently indicate symmetry lower than that of the internal structure, and (b) the conditions in which a crystal can exhibit higher symmetry than that corresponding to its structure. (We are not here concerned with the development of special forms (poles lying on symmetry elements), which are not characteristic of a single point-group and which, taken alone, would indicate a false symmetry, but with cases where the number and/or size of faces of a general form apparently indicate point-symmetry higher than the structural symmetry.) A simple example is the bipyramidal, instead of polar, development of certain orthorhombic crystals in the class  $mm$ .



(a) *Morphological Symmetry lower than Structural Symmetry.*

The nitrates of lead and barium, and the anhydrous nitrates of strontium and calcium, are all isomorphous. The space-group is  $P\bar{a}3$

Fig. 14.



The surface structure of an octahedron of lead nitrate. The nitrateion shown are nearly coplanar with the lead ions (small open circles).

and the point-symmetry should accordingly be  $m\bar{3}$ . As the result of numerous morphological studies they are assigned to the point-group 23. In the point-groups 23 and  $m\bar{3}$  the forms are as follows :

	23	$m\bar{3}$
100	Cube	Cube
110	Rhombic dodecahedron	Rhombic dodecahedron
111	Tetrahedron	Octahedron
$h k 0$	Pentagonal dodecahedron	Pentagonal dodecahedron
$h k k$	Triakis tetrahedron	Icositetrahedron
$h h l$	Deltoid dodecahedron	Triakis octahedron
$h k l$	Tetrahedral pentagonal dodecahedron	Dyakis dodecahedron

According to Groth {211} on barium nitrate is a triakis tetrahedron and {421} a tetrahedral pentagonal dodecahedron. (It is also stated that {111} is often present as unequally developed positive and negative tetrahedra, rather than as an octahedron, and also that there are often differences in the superficial appearance of {111} and  $\{1\bar{1}1\}$ . Crystals of these nitrates are not, however, optically active.

Now, although the normal procedure in deducing point-symmetry is to attach most importance to the number and arrangement of the faces of the more complex forms, we shall adopt a different viewpoint. In the case of these nitrates the equilibrium form is certainly  $\{111\}$ , and the way in which these faces develop should enable us to distinguish between the point-groups  $23$  and  $m3$ . In other words, we attach more importance to the development of these  $(111)$  faces than to the development of, say,  $\{211\}$  or  $\{421\}$ , on the grounds that  $\{111\}$  is the most stable form and  $\{211\}$  and  $\{421\}$  do not occur on an equilibrated crystal. In discussing the morphology of barium and lead nitrates it is convenient to compare them with potassium alum and sodium chlorate, for alum has the same space-group and its morphology is normal ( $m3$ ), while sodium chlorate is a normal example of a crystal with point-symmetry  $23$  (space-group  $P 2_13$ ), thus :

	Point-symmetry	
	From face development	From space-group
Alum .. ..	$m3$	$m3$
Barium nitrate ..	$23$	$m3$
Sodium chlorate ..	$23$	$23$

We have already discussed some of the possible reasons why a tetrahedral rather than an octahedral development might occur. Here it is only necessary to add a few experimental facts about the development of  $\{111\}$  on crystals of these nitrates.

(i.) Typical crystals of lead nitrate, alum and sodium chlorate obtained by cooling drops of supersaturated solution on a microscope slide are shown in Pl. XIV. (a)–(c). (Since sodium chlorate crystallizes from water as cubes, a little thiosulphate was added to the solution to produce  $(111)$  faces.) It is seen that the crystals of lead nitrate are mostly flattened octahedra very much like those of alum, though a few triangular crystals are always observed (see Section 5). The crystals of sodium chlorate, on the other hand, are always unmistakably tetrahedral in character. For comparison with the latter, Pl. XIV. (d) shows crystals of lead nitrate grown in the presence of a little methylene blue (which encourages  $\{100\}$ ). There is no preferential truncation of two of the four upper cube corners as in the case of sodium chlorate.

(ii.) Small, slowly grown, crystals of barium (or lead) nitrate are usually nearly perfect octahedra. Continued growth of a crystal resting on one  $(111)$  face leads to a rather flat crystal of which the three upper corners become truncated by cube faces (fig. 7 (b)). Octahedra of alum grow in exactly the same way if left resting on one  $(111)$  face. We have seen in Section 5 that it is not possible to obtain useful evidence as to the tetrahedral or octahedral nature of crystals grown in this way. However, a few crystals in any batch grow resting on a  $(100)$  face, and these develop

four equal (111) faces. Perfectly symmetrical "half-octahedra," fig. 5 (a), have been grown by mounting such a crystal on a cover-slip and floating it on the surface of a solution so that the crystal grows downwards into the solution with a four-fold axis vertical (see Pl. I.). These observations show that the morphology of these nitrates resembles much more closely that of alum than that of sodium chlorate, for if conditions are suitable the development is octahedral, not tetrahedral.

(b) *The "Holohedral" Developments of Hemihedral Crystals.*

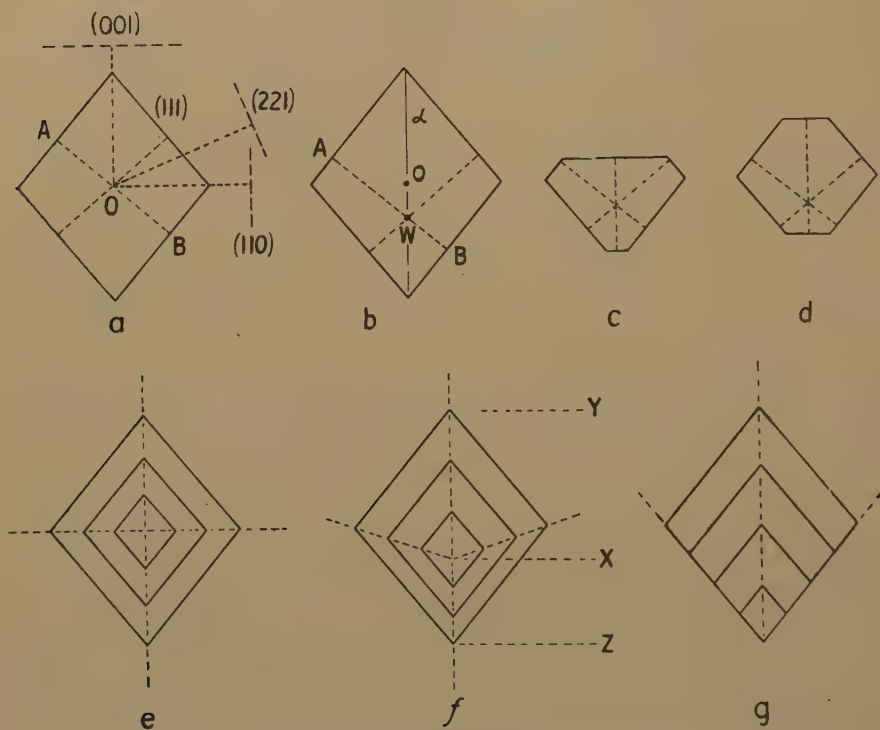
In order to avoid generalities we shall confine our discussion to certain crystals in the orthorhombic class  $mm$ . The form  $\{hkl\}$  consists of four faces only, forming a pyramid, and  $\{0kl\}$  of two faces. Anthranilic acid (space-group  $P2cn$ ) crystallizes from ethyl alcohol, ethyl acetate and aniline as bipyramids  $\{211\}$ , from water, glacial acetic acid and nitrobenzene as prismatic crystals,  $\{100\}$ ,  $\{010\}$  and  $\{211\}$ . In all cases the face development is the same at both ends of the polar axis, although the crystals are pyroelectric and undoubtedly belong to the class  $mm$ .  $\alpha$ -resorcinol (space-group  $Pna$ ) crystallizes from water as prisms  $\{110\}$  terminated symmetrically at both ends of the polar  $c$  axis by  $\{011\}$ , *i. e.* (011) and (0 $\bar{1}$ 1) develop similarly to (0 $\bar{1}$ 1) and (01 $\bar{1}$ ). From certain other solvents, a polar development is observed, *e. g.*  $\{011\}$  at one end and  $\{11\bar{1}\}$  at the other.

This question of polar or symmetrical development is best discussed in terms of the position of the Wulff point. For a holohedral crystal (class  $mmm$ ) the Wulff point is at the geometrical centre of the crystal (fig. 15 (a)). Let us assume that the equilibrium form is the bipyramid  $\{111\}$ , *i. e.* the surface free energies of all other faces are such that these faces lie outside the bipyramid and do not occur. In fig. 15 (a)  $OA=OB=k \cdot g_{111}$ . In the class  $mm$ ,  $g_{111} \neq g_{\bar{1}\bar{1}\bar{1}}$ , and the Wulff point is displaced (along the polar  $c$  axis) from the centre of the bipyramid, as in fig. 15 (b), and  $WA=k \cdot g_{111}$  and  $WB=k \cdot g_{\bar{1}\bar{1}\bar{1}}$ . The shape of the crystal is still necessarily a bipyramid. The hemihedral symmetry will be shown only if other faces with intercepts on the  $c$  axis appear, and then only in certain circumstances. For example, if  $g_{001} \approx g_{00\bar{1}}$ , the difference between the areas of (001) and (00 $\bar{1}$ ) is marked, fig. 15 (c), but if  $k \cdot g_{001} = k \cdot g_{00\bar{1}} + 2(OW)$ , a symmetrical development still results (d). Since  $2OW = k \cdot \text{cosec } \alpha (g_{111} - g_{\bar{1}\bar{1}\bar{1}})$ , the condition is  $g_{001} - g_{00\bar{1}} = \text{cosec } \alpha (g_{111} - g_{\bar{1}\bar{1}\bar{1}})$ . Thus we have the result, at first sight paradoxical, that the presence of (001) and (00 $\bar{1}$ ) with equal areas means, in this case, not that  $g_{001}$  and  $g_{00\bar{1}}$  are equal but that they are related as shown above.

Now, although there is obviously no direct way of locating the Wulff point, its position can be determined indirectly as follows:—A crystal of type (a) grows symmetrically, (e), but a crystal of type (b) unsymmetrically, (f), the rates of deposition on (111) and ( $\bar{1}\bar{1}\bar{1}$ ) being different. This can be detected if we can record the life-history of the crystal. A solution of anthranilic acid, initially colourless, darkens on keeping owing to the

formation of coloured oxidation products. The solubilities of these products relative to that of anthranilic acid are different in different solvents, with the result that from certain solvents (*e.g.* glacial acetic acid) the crystals are colourless, while from others (water, ethyl alcohol) they are coloured to various degrees. By growing a large crystal from a small coloured seed crystal the different extents of the growth in different directions can be observed relative to the Wulff point, which we know to be within the original seed crystal. In this way the ratio  $g_{111}/g_{\bar{1}\bar{1}\bar{1}}$  can be determined as  $XY/XZ$  in fig. 15 (*f*). Since the ratio of these two surface

Fig. 15.



The growth of orthorhombic crystals (see text).

free energies may vary with the nature of the solvent, it is clear that in such a case the Wulff point is not fixed but its position is determined by the environment of the crystal. Inclusions in a crystal may also indicate the position of the Wulff point (see below and Pl. XIV.).

The limiting case is when the Wulff point is actually at one end of the crystal. This would imply that  $g_{\bar{1}\bar{1}\bar{1}}$  is negligible compared with  $g_{111}$ , and growth would take place in one direction only (fig. 15 (*g*)). This is the case with crystals of resorcinol grown from water and certain other solvents, the inclusions in which show that growth is of this type. The crystals are prisms  $\{110\}$  terminated at one end by  $(011)$  and  $(0\bar{1}\bar{1})$  and at the other by  $(0\bar{1}\bar{1})$  and  $(011)$ . Owing to the way in which the molecules



are oriented the two former faces are hydroxylic and the latter benzenoid in character. The powerful interaction of the hydroxyl groups with the solution reduces the surface free energy of the former to a negligible fraction of that of the latter. The Wulff point, therefore, lies in the edge between (011) and (0 $\bar{1}$ 1) and no growth occurs on these faces. The morphology of resorcinol will be discussed in more detail in a later paper.

We see, therefore, that this concept of the Wulff point is of great value in considering the face developments of crystals. Although in certain cases pairs of parallel pyramid faces may be structurally non-equivalent, this is not evident from the geometry of the bipyramid, and observations on the mode of growth of the crystal are necessary to distinguish the various cases. The Wulff point is sometimes at the geometrical centre of the crystal, in other cases it is in less symmetrical positions which may vary according to the nature of the solvent. In the limiting case it may be in a surface (edge or apex) of the crystal, as when the crystal is growing from a surface, when there is very strong interaction between the solution and certain faces of the crystal, or when there is very strong adsorption of a foreign material.

#### SUMMARY.

The Gibbs postulate, that the equilibrium shape of a crystal is that of minimum total surface free energy for a given volume, is used in a discussion of some topics in morphological crystallography. The theorem of Wulff is applied in considering the effect of environment on crystal habit, some cases of anomalous face development, and the conditions which must be satisfied in order that measurements of the relative rates of growth normal to crystal faces can yield ratios of surface free energies. It is asserted that the equilibrium shape of a crystal is likely to be a relatively simple development, and some of the causes of more complex development are discussed. The complexity of face development which is a striking feature of the descriptions of so many crystals, both natural and artificial, gives an entirely misleading picture of the fundamental morphology of the equilibrated crystals.

#### DESCRIPTION OF PLATE XIV.

(a)–(d) Photomicrographs of lead nitrate, potassium alum, sodium chlorate (from solution containing thiosulphate), and lead nitrate (from solution containing methylene blue) respectively.

(e) Crystal of sodium chlorate from solution containing thiosulphate, showing truncation of only two upper corners.

(f) Crystal of lead nitrate showing symmetrical development of four octahedron faces.

(g) Half-tetrahedra of sodium chlorate.

(h) "Rhombohedral" crystal of barium nitrate (six {111} faces).

(i) and (j) Symmetrical inclusions in crystals of catechol and thiourea, respectively.

(k) Unsymmetrical inclusions in a crystal of resorcinol from aqueous solution.

(l) Colourless crystal of anthranilic acid grown from coloured seed-crystal.

*References.*

- Donnay and Harker, 'Naturaliste Canadien,' lxvii. p. 33 (1940).  
 France, J. Phys. Chem. xxxix. p. 425 (1935).  
 Kolb and Comer, J.A.C.S. lxvii. p. 895 (1945).  
 Laue, *Z. Krist.* cv. p. 124 (1943).  
 Miers, Phil. Trans. ccii. p. 459 (1904).  
 Neuhaus, *Z. Krist.* lxviii. p. 15 (1928).  
 Spangenberg, *ibid.* lix. p. 375 (1924).  
 Tremblay, J. Wash. Acad. Sci. xxxii. p. 327 (1942).  
 Wulff, *Z. Krist.* xxxiv. p. 449 (1901).

Imperial Chemical Industries, Ltd.,  
 Research Laboratories,  
 Blackley, Manchester, 9.

LXII. *The Ionization of Mercury Vapour by Slow Electrons.*

By the late T. McFADDEN, Ph.D., Physics Department,  
 Queen's University, Belfast\*.

[Received August 13, 1946.]

The experiments described in this paper were commenced by Dr. McFadden shortly before the war and continued subsequently as part of work done in this Department for the Director of Scientific Research, Admiralty. We are now indebted to the Board of Admiralty for permission to publish the paper, which has been prepared from Dr. McFadden's notes with the help of Dr. R. H. Sloane. Tube A was built for the advanced undergraduate laboratory by Dr. Sloane, and the existence of interesting details in curves obtained with it first noticed when it was in use for this purpose.—K. G. E.

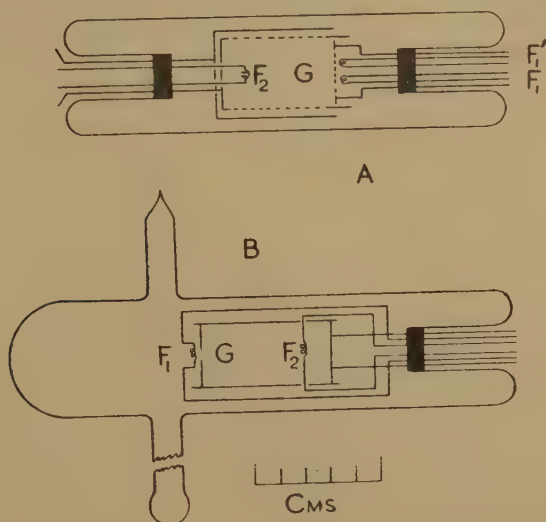
It has been shown by W. B. Nottingham, and others<sup>(1)</sup>, that the chance of ionization of mercury by electron impact does not increase regularly with electron energy immediately above the first ionization potential, 10.38 volts. Instead, it passes through a maximum at 10.8 volts, a minimum at 11.05 volts, a second, less pronounced maximum at 11.45 volts, and thereafter, although probably continually increasing up to at least 12.5 volts, does so irregularly. These fluctuations are of interest in connection with the theory of atomic collisions, and are possibly of importance technically in hot-cathode mercury vapour rectifying tubes and thyratrons. Although Nottingham's experiments were done with great care, with electrons made closely homogeneous in speed by magnetic sorting, it appears therefore worth reporting some preliminary experiments, which have been done on this problem with a space-charge detector, used both in the conventional manner and in a sensitive modified arrangement. The results which have been obtained are in the main

\* Communicated by Prof. K. G. Emeleus.

consistent with those reported by Nottingham, and, in addition, raise some points of technique and interpretation in the measurement of critical potentials.

The two tubes used are shown in fig. 1. Most of the measurements

Fig. 1.



were made with tube A. When using this, electrons were drawn from one of two similar filaments,  $F_1$  and  $F_1'$ , through a wire window of wide mesh (0.1 mm. diameter wire woven across a hole in a zigzag with the end spacings about 0.5 mm. apart), into an equipotential box G, which contained a second filament  $F_2$  to act as a space-charge detector of positive ions. The positive end of  $F_2$  was at the same potential as G. A few experiments were done with tube B, which had a slightly different electrode structure, electrons entering G through an unbridged round hole 2 mm. in diameter. In both tubes all metal parts except the filament were made of nickel or nichrome. The tubes were sealed off from the pumps, and each contained a pellet of mercury, which was kept as far as possible in a side tube for control of vapour pressure. In most cases the side tube was held at 22° C., but in a few cases it was at 36° C., in each case with the rest of the tube at a higher temperature. It is difficult to be sure that the vapour pressures were not sometimes higher than the saturation pressures at these temperatures ( $1 \cdot 10^{-3}$  mm. and  $4 \cdot 10^{-3}$  mm. respectively) owing to temporary formation of drops in the electrode structure near the filaments. The magnitude of the ionization measured by  $F_2$  above 10.4 volts in tube A, using the a.c. method (see below), was, however, found in some control experiments to be nearly proportional to the current from  $F_1$ , showing that it was due largely to single impact and not to cumulative processes. The tubes, especially B, were prepared with

precautions approaching those described by Nottingham, but without use of a barium getter, and in the case of tube A, with less careful distillation of the mercury.

Some measurements were made with both tubes by the straightforward Hertz method<sup>(2)</sup>. Changes in the current  $i_2$  from the space-charge filament  $F_2$ , whose main component through the galvanometer recording it was balanced out, were measured, as a function of the accelerating voltage  $V$  between  $F_1$  or  $F_1'$  and  $G$ . Typical results are shown in figs. 2 and 3. The voltage scale was decided on after correlation of the data

Fig. 2.



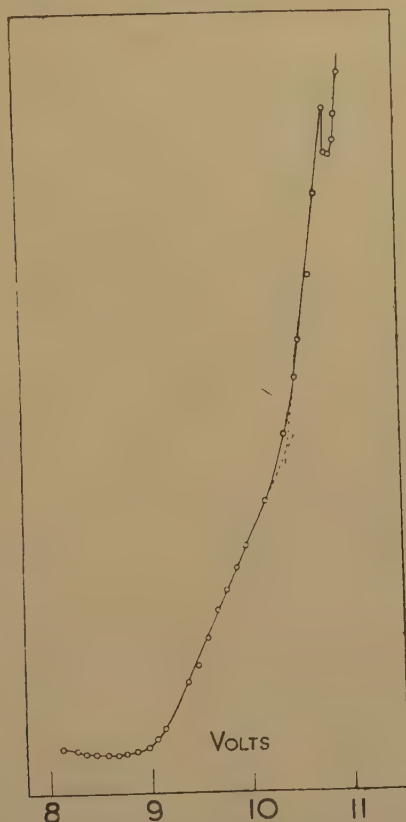
Ionization below 10.4 volts. Ordinates change of space-charge detector current ( $i_2$ ): abscissæ, accelerating voltage  $V$ . The broken line was covered by numerous experimental points not shown. Temperature, 22°C.



with those of Arnot and Milligan<sup>(3)</sup> and of Nottingham<sup>(1)</sup>, and after consideration of the various critical potentials shown by these and other curves. The zero correction necessary relative to the measured voltage ranged from 0.2 to 0.9 volts. The two principal features of these curves are :—

- (1) They provide evidence for the existence under these conditions of considerable ionization below 10.4 volts, almost certainly molecular ( $\text{Hg}_2$ ) in origin.

Fig. 3.

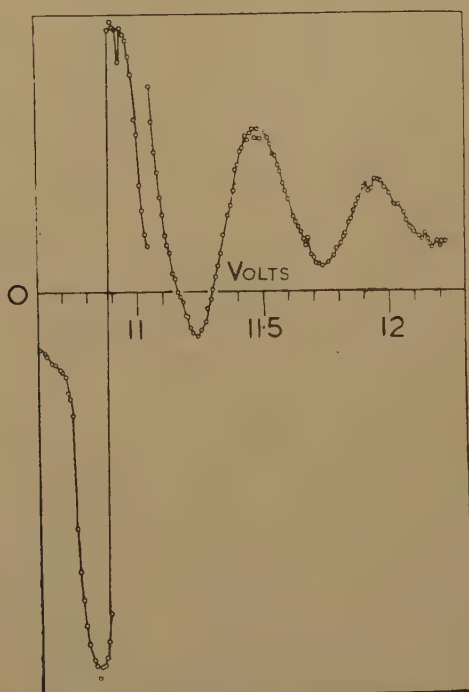


- (2) Fig. 3 indicates that there is a minimum in the ionization cross-section  $v$  voltage curve at about 10.85 volts. This has been recorded repeatedly, and is believed to be the same as that recorded by Nottingham<sup>(1)</sup> in the same voltage region.

Tube A has also been studied by a method similar to one previously used in the analysis of the collector characteristic curves for an exploring electrode in a plasma<sup>(4)</sup>. The aim in the present case is to bring out details in the curve giving  $i_2$  as a function of  $V$  in the atomic ionization range, by a direct measurement of  $d^2i_2/dV^2$ , which will be referred to as  $D_2$ . This is accomplished, to a constant factor, by measuring the change

in  $i_2$  when  $V$  is modulated by a small alternating potential  $a \sin 2\pi nt$  from an audio-frequency oscillator. The change is nearly proportional to  $a^2 D_2$ . It is assumed there is no difference between the static and dynamic  $i_2 v V$  characteristics. The value of  $n$  was varied from 1060 to 1600 per second without noticeable effect on the results, and that of  $a$  from 0.002 volts to 0.02 volts without effect other than the consequent increase in  $i_2$ . A typical  $D_2 v V$  curve is shown in fig. 4. The voltage scale has been decided on by comparison with curves like those of fig. 3, but is unfortunately uncertain to about 0.4 volt. This is because it was not found practicable to take data for curves like those of figs. 3 and 4

Fig. 4.



Second differential curve ( $D_2$ ) of  $i_2 v V$ . The reality of the small dip at the first maximum, and of the other small irregularities, is doubtful. Above 11.05 volts the scale of  $D_2$  is increased by a factor of 10. The horizontal line is the zero of  $D_2$ . Temperature  $22^\circ \text{C}$ .

simultaneously at the stage to which the work was developed. In repeated experiments curves like that of fig. 4 were, however, obtained consistently; the first large double peak was occasionally sharper. The positions of the maxima and minima at higher voltages could not be fixed definitely relative to that of the first maximum on this account, but were concordant to about 0.2 volt or better. Some small secondary peaks on the main waves mentioned in an earlier note <sup>(5)</sup> are now believed to be spurious.

The obvious interpretation of the irregularity in fig. 3 at about 10.8 volts, and of the waves in fig. 4, is that they are due to the same

causes as irregularities in the ionization cross-section curves found by Nottingham <sup>(1)</sup>, *i. e.* that they are of atomic origin. The detail obtained would require that the effective energy spread of the beam from  $F_1$  was decidedly less than 0.2 volt. This is reasonable, as the fall in potential across the filament was about 0.4 volt, and the main emission probably came from one or two small turns in the middle. An exact comparison between Nottingham's ionization-probability  $v$  voltage curves and fig. 4 is scarcely practicable, as the former are jagged, making the determination of curvature difficult. The main features appear similar, however, at least up to about 11.6 volts. Exact agreement on the voltage scale might indeed not be expected, as it is conceivable that the contact p.d. between  $F_1$  and G (fig. 1) depends on the rate of arrival of positive ions at  $F_1$  from the space between  $F_1$  and G, which will vary with  $V$ . The electron current from  $F_1$ , which was nearly saturated, showed minute fluctuations of order of 0.5 per cent. or less of the total current, between about 10.5 and 12 volts, at about the same voltages as the waves of fig. 5. An arc did not strike between  $F_1$  and G below 13 volts. The field from  $F_1$  therefore extended to G in the present experiments, and was not concentrated in a thin double sheath on  $F_1$ , although the space-charge between  $F_1$  and G may have included a positive section. It is assumed, as in some earlier investigations <sup>(6)</sup>, that these fluctuations in current from  $F_1$  are a consequence of the change in probability of ionization with voltage.

It is also assumed that the waves of fig. 4 are not produced by geometrical periodicities in the tube. There are three obvious periodicities: (1) the turns of  $F_1$ , (2) the wires through which the electrons enter G, (3) the turns of  $F_2$ . The electrons able to produce atomic ionization in G will form a somewhat laterally restricted beam when  $V$  is very slightly greater than 10.4 volts. As  $V$  increases and more positive parts of  $F_1$  contribute electrons able to ionize, the effective width of the beam will increase. It might be possible for the width to increase irregularly on account of either of the two first periodicities, and for this to give rise to an apparently wavy ionization function. Again, even if the rate of increase of width of the beam with  $V$  were steady, waves might result if the beam swept across the turns of  $F_2$ . The equivalent of a geometrical periodicity would also result if the thermionic emission constants of the material of  $F_1$  varied along its length. Further experiments are required to settle these points, and the exact position of the voltage scale of fig. 4. Until these have been done the possibility that the first wave of fig. 4 marks the onset of atomic ionization at 10.4 volts cannot be ruled out.

### References.

- (1) Nottingham, *Phys. Rev.* lv. p. 203 (1939).
- (2) Hertz, *Zeitschrift für Physik*, xviii. p. 307 (1923).
- (3) Arnot and Milligan, *Proc. Roy. Soc. A*, cliii. p. 359 (1936).
- (4) Sloane and Macgregor, *Phil. Mag.* xviii. p. 193 (1934).
- (5) McFadden, *Phys. Rev.* lv. p. 797 (1939).
- (6) *E. g.* Benade and K. T. Compton, *Phys. Rev.* xi. p. 184 (1918).

LXIII. *On the Radial Error in a Gaussian Elliptical Scatter.*

By D. MARTIN, M.A., B.Sc.\*

[Received October 18, 1944.]

*Summary.*

Considering a bivariate normal distribution of errors, expressions are derived for

- (a) the probability that the radial error has less than any assigned value, and
- (b) the mean radial error.

1. If an infinite number of particles are scattered about the origin and in the  $xy$  plane according to the Gaussian normal law (*e.g.* shots aimed at the origin by a rifleman), then the probability of a particle being inside an element of area  $dxdy$ , situated at the point  $(x, y)$ , is  $\phi(x, y)dxdy$ , where

$$\phi(x, y) = \frac{hk}{\pi\sqrt{1-\rho^2}} \exp[-(h^2x^2 - 2\rho h kxy + k^2y^2)/(1-\rho^2)]; \quad (1.1)$$

$h, k$  are the moduli of precision and  $\rho$  is the coefficient of correlation.

It follows from (1.1) that the loci of equal probability density constitute a family of ellipses homothetic with respect to the origin, and it is not difficult to calculate the probability of a particle being found inside any particular one of these ellipses. However, the corresponding problem for a circle with centre at the origin and any assigned radius is more difficult, and so far does not seem to have been dealt with. An attempt to solve the problem is made in the following lines.

2. Assuming for the moment that the variates are uncorrelated, the probability  $P$  of a particle lying within a radius  $R$  of the origin is given by

$$P = \frac{hk}{\pi} \iint e^{-\frac{1}{2}(h^2x^2 + k^2y^2)} dxdy$$

over the circle  $x^2 + y^2 = R^2$  or, on changing to polar co-ordinates  $(r, \theta)$  and putting  $\phi = 2\theta$ , by

$$P = \frac{hk}{\pi} \int_0^R \int_0^{2\pi} r e^{-\frac{1}{2}r^2\{h^2 + k^2 + (h^2 - k^2)\cos\phi\}} d\phi dr. \quad \dots (2.1)$$

Now

$$\int_0^{2\pi} e^{\pm x \cos \phi} d\phi = 2\pi I_0(x),$$

---

\* Communicated by Professor T. M. MacRobert.



where  $I_0(x)$  is the Modified Bessel Function of the First Kind and Zero Order; hence

$$P = 2hk \int_0^R e^{-\frac{1}{2}r^2(h^2+k^2)} I_0\left\{\frac{1}{2}r^2(h^2-k^2)\right\} r dr.$$

On changing the variable, and putting

$$\lambda = R^2 h^2, \quad k/h = s, \quad \dots \dots \dots (2.2)$$

this equation becomes

$$P = \frac{2s}{s^2-1} \int_0^{\frac{1}{2}\lambda(s^2-1)} e^{-\frac{s^2+1}{s^2-1}x} I_0(x) dx. \quad \dots \dots \dots (2.3)$$

$s$  can be considered as lying either between 0 and 1, or between 1 and  $\infty$ ; for definiteness, we take the latter alternative (the major axes of the ellipses of equal probability density then lie along the  $x$ -axis).

The alternative representation, obtained by changing the order of integration in (2.1), is

$$P = 1 - \frac{hk}{\pi} e^{-\frac{1}{2}R^2(k^2+h^2)} \int_0^{2\pi} \frac{e^{\frac{1}{2}R^2(k^2-h^2)\cos\phi}}{k^2+h^2-(k^2-h^2)\cos\phi} d\phi \quad \dots \dots (2.4)$$

but, for the present purpose, this appears to be less useful than (2.3).

Reverting to (2.3), the integral to be evaluated is of the form

$$u = \int_0^\beta e^{-x \cosh \alpha} I_0(x) dx, \quad \alpha, \beta \text{ real.} \quad \dots \dots \dots (2.5)$$

The integration can be carried out by means of a device used by Erdélyi (1939) for dealing with a similar integral.

We begin with Jacobi's formula (Watson (1922), p. 22),

$$e^{-ix \cos \alpha} = J_0(x) + 2 \sum_{n=1}^{\infty} i^{-n} J_n(x) \cos n\alpha, \quad \dots \dots \dots (2.6)$$

where  $J_n(x)$  denotes the Bessel Function of the First Kind and Order  $n$ . The expansion is uniformly convergent for all finite values of  $x$  and  $\alpha$ , real and complex. Thus, since

$$I_n(x) = i^{-n} J_n(ix), \quad \dots \dots \dots (2.7)$$

it follows that

$$e^{(\beta-x) \cosh \alpha} = I_0(\beta-x) + 2 \sum_{n=1}^{\infty} I_n(\beta-x) \cosh n\alpha.$$

Differentiation of both sides of this equation with respect to  $\alpha$  yields

$$e^{(\beta-x) \cosh \alpha} = \frac{2}{\sinh \alpha} \sum_{n=1}^{\infty} \frac{n \sinh n\alpha I_n(\beta-x)}{\beta-x},$$

the process being legitimate by reason of the uniform convergence of the series obtained.

Reverting now to (2.5),

$$u = \frac{2e^{-\beta \cosh \alpha}}{\sinh \alpha} \int_0^\beta \sum_{n=1}^{\infty} \frac{n \sinh n\alpha I_n(\beta-x) I_0(x)}{\beta-x} dx. \quad \dots \dots (2.8)$$

Now Kapteyn's Integral (Watson (1922), p. 380),

$$\int_0^\beta J_\nu(x) J_n(\beta-x) \frac{dx}{\beta-x} = \frac{1}{n} J_{n+\nu}(\beta),$$

remains valid when the ordinary Bessel functions are replaced by modified functions throughout. Term by term integration of (2.8) is legitimate, and so

$$u = \frac{2e^{-\beta \cosh \alpha}}{\sinh \alpha} \sum_{n=1}^{\infty} I_n(\beta) \sinh n\alpha. \quad (2.9)$$

Hence, putting  $\cosh \alpha = (s^2+1)/(s^2-1)$  and  $\beta = \frac{1}{2}\lambda(s^2-1)$ , (2.2), (2.3), (2.5) yield

$$P = e^{-\frac{1}{4}R^2(k^2+h^2)} \sum_{n=1}^{\infty} \left\{ \left( \frac{k+h}{k-h} \right)^n - \left( \frac{k+h}{k-h} \right)^{-n} \right\} I_n \left\{ \frac{1}{2}R^2(k^2-h^2) \right\}. \quad (2.10)$$

When  $n$  is large,

$$\frac{I_{n+1} \left\{ \frac{1}{2}R^2(k^2-h^2) \right\}}{I_n \left\{ \frac{1}{2}R^2(k^2-h^2) \right\}} \sim \frac{R^2(k^2-h^2)}{4(n+1)},$$

whence it follows that the series (2.10) converges at a rate comparable with that of the series for  $\exp \left\{ \frac{1}{4}R^2(k+h)^2 \right\}$ , and is therefore suitable for numerical applications if  $R$  is small. Bessel functions are extensively tabulated but, if a few of the functions required are of a higher order than those available, they can be readily computed from the recurrence formula

$$I_{n+1}(x) = I_{n-1}(x) - \frac{2n}{x} I_n(x). \quad (2.11)$$

Considering now the general case of  $\rho \neq 0$ ,

$$P = \frac{hk}{\pi \sqrt{(1-\rho^2)}} \iint \exp [-(h^2x^2 - 2\rho h k x y + k^2y^2)/(1-\rho^2)] dx dy$$

over the circle  $x^2 + y^2 = R^2$ .

By rotating the co-ordinate axes until they coincide with the axes of the homothetic ellipses, this integral reduces to the one already considered in the case  $\rho = 0$ . Thus, we find that (2.10) still holds provided that  $h$  and  $k$  are replaced by  $h'$  and  $k'$  respectively, where

$$\begin{aligned} h'^2 &= \frac{1}{2}[k^2 + h^2 - \{(k^2 - h^2)^2 + 4\rho^2 h^2 k^2\}^{\frac{1}{2}}] (1-\rho^2)^{-1}, \\ k'^2 &= \frac{1}{2}[k^2 + h^2 + \{(k^2 - h^2)^2 + 4\rho^2 h^2 k^2\}^{\frac{1}{2}}] (1-\rho^2)^{-1}, \end{aligned} \quad (2.12)$$

and the positive roots are taken.

3. We conclude this note by deriving a very simple expression for the mean distance of the particles from the origin (*i. e.* the mean radial error).

If  $\rho = 0$ , the mean radial error  $\bar{E}_r$  is given by

$$\bar{E}_r = \frac{hk}{\pi} \int_0^\infty \int_0^{2\pi} r^2 e^{-\frac{1}{4}r^2 \{k^2 + h^2 - (k^2 - h^2) \cos \phi\}} d\phi dr; \quad (3.1)$$

\* This result seems to be capable of further simplification in only one case, *viz.*,  $\alpha = 0$ . Using (2.11) it easily follows that

$$\int_0^\beta e^{-x} I_0(x) dx = \beta e^{-\beta} \{I_0(\beta) + I_1(\beta)\},$$

a result which has already been noted by McLachlan and Meyers (1936).

the derivation of (3.1) is similar to that of (2.1). Hence, integrating with respect to  $r$ ,

$$\bar{E}_r = \frac{hk}{\sqrt{2\pi}} \int_0^{2\pi} \frac{d\phi}{\{k^2 + h^2 - (k^2 - h^2) \cos \phi\}^{3/2}}. \quad (3.2)$$

Consider the integral

$$u = \int_0^\pi \frac{d\phi}{(1 - \alpha \cos \phi)^{3/2}} = \int_0^\pi \frac{d\phi}{(1 + \alpha \cos \phi)^{3/2}}, \quad 0 \leq \alpha \leq 1;$$

on putting  $\phi = 2x$ , this becomes

$$u = \frac{2}{(1 + \alpha)^{3/2}} \int_0^{\pi/2} \frac{dx}{(1 - \frac{2\alpha}{1 + \alpha} \sin^2 x)^{3/2}},$$

and reduces to

$$u = \frac{2}{(1 - \alpha)\sqrt{\alpha + 1}} E\left(\sqrt{\frac{2\alpha}{1 + \alpha}}\right),$$

where  $E(\kappa)$  is Legendre's Complete Elliptic Integral of the Second Kind and modulus  $\kappa$ . The integral is tabulated against the modulus in various books, *e. g.* Jahnke-Emde (1938).

Hence (3.2) becomes

$$\bar{E}_r = \frac{1}{h\sqrt{\pi}} E\left(\sqrt{1 - \frac{h^2}{k^2}}\right). \quad (3.3)$$

When  $\rho \neq 0$ , the same result holds provided that  $h, k$  are replaced as before by  $h', k'$ , respectively.

*Note added in proof.*—I have recently come across a paper on ærodynamics (v. Bonbély, *Z. Angew. Math. Mech.* xxii. p. 190 (1942) in which the expansion

$$J_0(u \sin \theta) = \sum_{n=0}^{\infty} \frac{(1 - \sin^2 \theta)^n u^n}{2^n n!} J_n(u)$$

is used to derive a series expansion for

$$\int_0^\omega e^{-iu} J_0(u \sin \theta) d\theta.$$

The analogous result for (2.5) leads to

$$P = R^2 h k e^{-\frac{1}{2} R^2 (k^2 + h^2)} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! (2n+1)} \left( \frac{R^2 h^2 k^2}{k^2 + h^2} \right)^n \\ \times [I_n\{\frac{1}{2} R^2 (k^2 + h^2)\} + I_{n+1}\{\frac{1}{2} R^2 (k^2 + h^2)\}].$$

This series is more rapidly convergent than (2.10).

#### REFERENCES.

- Erdélyi, *Proc. Edinb. Math. Soc.* (2) vi. 94 (1939).  
 Jahnke-Emde, 'Tables of Functions', 3rd ed. (1938).  
 McLachlan and Meyers, *Phil. Mag.* xxi. p. 444 (1936).  
 Watson, 'Treatise on Bessel Functions' (1922).

LXIV. *Scattering of X-rays of High Frequency.*

By N. WINOGRADOFF, Ph.D., A.Inst.P.\*

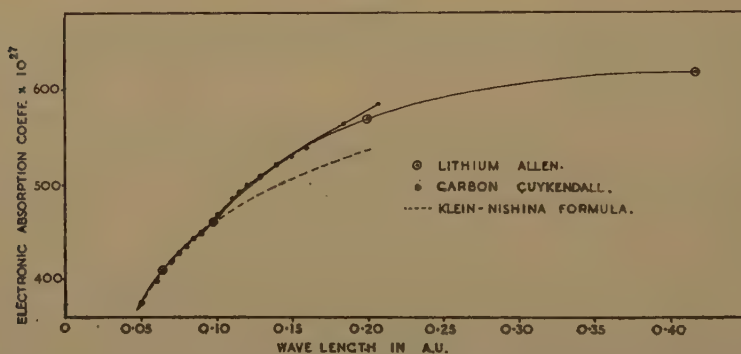
[Received December 19, 1945.]

In investigating experimentally the scattering of X-rays of high frequency by elements of low atomic number, the general procedure is to measure the electronic absorption coefficient,  $\mu_e$ , of the radiation and assume that if very short wave-lengths are used, the photo-electric absorption is negligible in comparison with the absorption due to scattering, so that  $\mu_e$  may be taken to represent the electronic scattering coefficient.

The results of this type of experiment reported by Allen<sup>(1)</sup>, Cuykendall<sup>(2)</sup>, and others have shown that for scatterers of low atomic number ( $N < 8$ ) the electronic absorption coefficients,  $\mu_e$ , for wave-lengths below and up to about 0.10 A.U. were in very good agreement with the values predicted by the Klein-Nishina formula. For wave-lengths above about 0.10 A.U., however, it appeared that all the experimental values of the electronic absorption coefficients were considerably higher than the corresponding Klein-Nishina values.

This departure from the theoretical values is shown for carbon (Cuykendall), and lithium (Allen) in fig. 1. There appears to be a marked and

Fig. 1.



well-defined point of inflection in the experimental curves in the neighbourhood of  $\lambda = 0.10$  A.U.

Cuykendall suggested that at this and higher wave-lengths the photo-electric absorption becomes appreciable; this, however, can hardly be the correct explanation, since the excess absorption, *i. e.* the difference between the experimental and the Klein-Nishina values, denoted by  $\delta(\mu_e)$ , would then be expected to follow the  $\lambda^3$  absorption law which appears to hold between the characteristic absorption edges in the longer wave-length regions. When, however, the difference between the two values

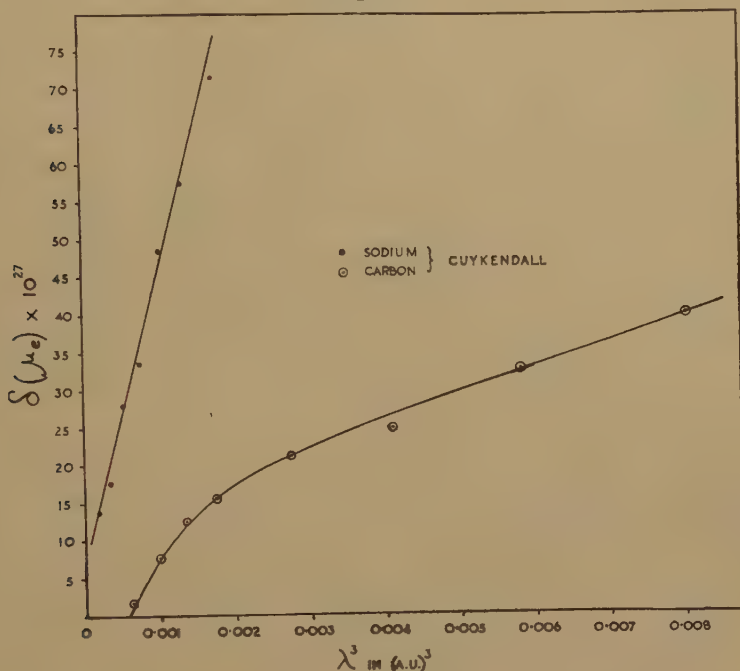
\* Communicated by the Author.



obtained for carbon is plotted against  $\lambda^3$  we get the lower curve shown in fig. 2, similar curves are obtained for lithium and aluminium; these are not straight lines at the shorter wave-lengths. On the other hand, if the same differences are plotted against  $1/\lambda$  the curves are straight lines at the shorter wave-lengths (fig. 3).

If we denote the difference between the experimental and Klein-Nishina values of the electronic absorption coefficients by  $\delta(\mu_e)$ , then the

Fig. 2.



excess absorption at these short wave-lengths appears to be represented by an equation of the form :

$$\delta(\mu_e) = \alpha - \frac{\beta}{\lambda},$$

where  $\alpha$  and  $\beta$  are constants depending on the atomic number of the scatterer. This form of the relationship between  $\delta(\mu_e)$  and the wave-length was found to hold for lithium, carbon and aluminium in the wave-length ranges :

$$0.094 < \lambda < 0.15 \text{ A.U.}$$

$$0.083 < \lambda < 0.16 \text{ A.U.}$$

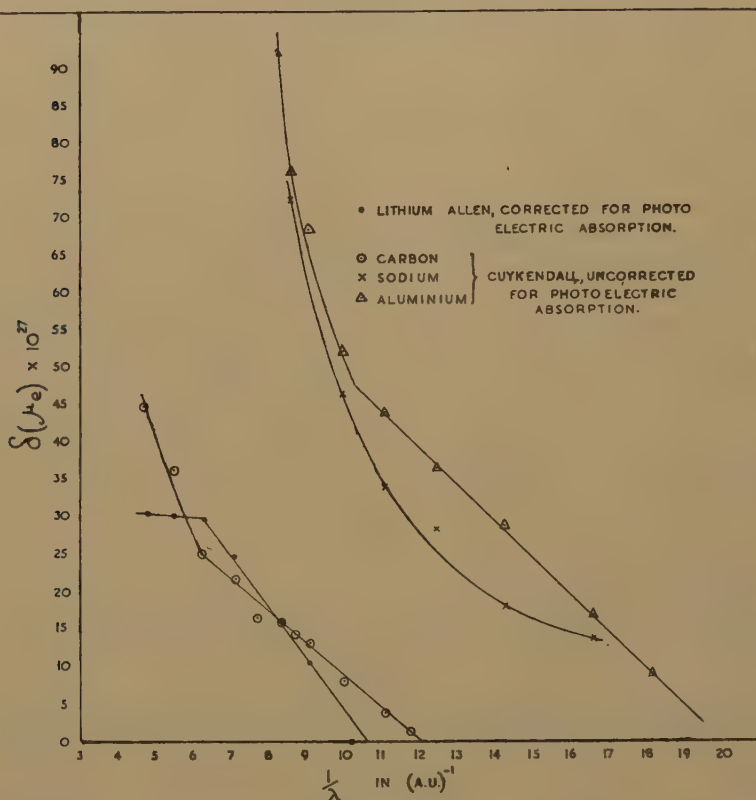
$$0.050 < \lambda < 0.096 \text{ A.U., respectively,}$$

but is not applicable to sodium, the latter apparently obeying the  $\lambda^3$  absorption law, as shown in fig. 2. (Experimental data for other elements of low atomic number are lacking.)

The above expression immediately suggests that the excess absorption at these wave-lengths cannot be explained in terms of the classical theory

of scattering, Compton's theory of scattering (represented by the Klein-Nishina formula) or by photo-electric absorption. In other words, at these wave-lengths there seems to be evidence of a new mechanism of absorption, which is associated with wave-lengths well below the characteristic K radiation of the scatterer. The theoretical interpretation of the phenomena should, therefore, prove to be particularly interesting and important.

Fig. 3.



It should be emphasized that the new mechanism of absorption appears to be responsible for only about seven per cent. of the total  $\mu_e$  at these wave-lengths, and would, therefore, require a very refined technique for experimental investigation, although in the experiments of Allen and Cuykendall the experimental error is only of the order of one or two per cent. The presence of this new mechanism of absorption may also be the cause of the apparent slight variations in the power of  $\lambda$  in the  $\lambda^3$  absorption law reported by many workers.

By a simple application of Heisenberg's Uncertainty Principle and relativistic mass corrections to the problem of scattering of radiation by a free electron, the writer has derived a new formula\* for the variation

\* It is hoped to give the derivation of this formula in a separate paper shortly.

of the electronic scattering coefficient  $\sigma_e$  with wave-length of the incident radiation; given by the expression

$$\sigma_e = \frac{8\pi}{3} \frac{e^4}{m_0^2 c^4} \left\{ 1 - \frac{4}{\pi} \frac{h\nu}{m_0 c^2} \right\},$$

where

$e$  is the electronic charge in e.s.u.

$m_0$  is the rest mass of the electron,

$c$  is the velocity of light.

$h$  is Plank's constant.

$\nu$  is the frequency of the incident radiation.

This formula gives values of  $\sigma_e$  agreeing to within one per cent. of the experimental values found by Allen for lithium in the wave-length range 0.098 A.U. to 0.4 A.U. and possibly for longer wave-lengths as well. The new formula thus actually gives values of  $\sigma_e$  which incorporate the excess absorption, mentioned above, and seems at present to indicate that the excess absorption should be attributed to a scattering rather than an absorption process.

Below 0.098 A.U. the values of  $\sigma_e$  given by the new formula fall short of the experimental values, and the difference plotted against the reciprocal of the corresponding wave-length is almost a straight line, *i. e.*

$$\delta(\sigma_e) = \alpha' - \frac{\beta'}{\lambda}.$$

It is interesting to note that J. M. Cork<sup>(3), (4)</sup> and Pidd<sup>(3)</sup> have recently shown that in this region the Klein-Nishina values are higher than the experimental values obtained for Cu, Fe, and Pb.

#### References.

- (1) Allen, Phys. Rev. xlv. p. 122 (1934).
- (2) Cuykendal, Phys. Rev. l. p. 105 (1936).
- (3) Cork and Pidd, Phys. Rev. lxvi. p. 227 (1944).
- (4) Cork, Phys. Rev. lxvii. p. 53 (1945).

#### LXV. An Illusion of Size.

By ROBERT WEIL, B.Sc., A.Inst.P.\*  
South-west Essex Technical College.

[Received October 17, 1945.]

PROF. LOEWENSTEIN reported recently † that, under suitable illumination, dull and bright nickel threepenny bits appeared to differ in size, the duller being the larger. A white background is said to invert the phenomenon, whereas a dark or neutral background seems to show it

\* Communicated by the Author.

† 'Nature,' clv. p. 672 (1945).

clearly. The following explanation is suggested: the old coin appears fogged, whereas the bright one has a sharp outline even in reduced light. Greater distances are associated with fogged objects, and there may result the tendency to over-estimate their size. The fact that the difference in size appears to be more telling after a few seconds is said to support an explanation on psychological grounds. The colour of the background in general seems to influence the magnitude of the effect in no way at all.

The matter was taken up by Prof. Hartridge\*, who found that old and new nickel threepenny bits actually differed in size. On choosing two equal coins which he placed on a black background, he found that the coins appeared to be either equal or the dull one was somewhat smaller, as might be expected owing to irradiation: these results were obtained with perpendicular illumination. But with oblique lighting the dull coin occasionally appeared to be the larger. The explanation advanced in this case was based on the assumption that the bright reflexion from the raised edge of the polished coin was smaller than the dull reflexion from the raised edge of the other; an interpretation of the phenomenon on psychological grounds is dismissed as improbable.

Mr. Phelps†, of the Royal Mint, suggests that a probable explanation might lie in the roundness of the edge of the coin so that the reflexion of narrow-angle illumination appeared a bright line, the outside of which would be considered as the edge of the coin, thus apparently reducing its size.

The present writer carried out a comprehensive set of experiments to test especially the "raised edge" theories. The influence of the absolute size of the object under observation was investigated, as were also the effect of varying the colour of the light and that of the background. An attempt was made to detect a difference in the manifestation of the effect with coins of different colours.

Parallel light from a carbon arc was used throughout the experiment, and its colour was varied by three Ilford filters and one neutral-tinted plate. The Ilford plates were Nos. 204, 304, 404, with approximately 60 per cent. transmissions for 6300 Å, to 7400 Å., 4300 Å., 5250 Å. respectively. The percentage is a maximum. The background consisted in turn of black, white, violet, blue, green, yellow, and red papers. The coins employed in the first experiment were farthings and pennies, silver threepenny bits and half-crowns. Oblique and perpendicular illumination was employed. The coins were viewed along the reflected beam in all instances. It was noted that diffused light made the dull coin appear larger in every case. The diameters of the paired coins were found to be equal when measured with ordinary engineering appliances. All the coins were originally dull, and one of each pair was polished repeatedly, particular care being taken to obtain the same brightness near the edge as at the centre. Heads were used as reflecting surfaces in all cases.

It is very cumbersome, of course, to carry out an experiment of such dimensions with, say, 50 people; where no agreement was obtained with

---

\* 'Nature,' clvi. p. 118 (1945).

† 'Nature,' clvi., p. 269 (1945).



three or four observers a question-mark was put down : O indicates that the dull, N that the bright coin appeared to be the larger of the two. A sign of equality represents instances where no difference could be detected. The second column stands for oblique, the third for perpendicular illumination : the first merely gives the sum of these.

TABLE I.

Results for the Individual Coins ; Effect of Size. The first column gives the sum of the second and third (oblique and perpendicular illumination respectively).

	Farthing.			Penny.			Silver threepenny bit.			Half-crown.		
O	12	6	6	2	1	1	32	16	16	8	1	7
=	19	12	7	23	14	9	16	9	7	37	17	20
N	23	10	13	29	12	17	6	2	4	11	10	1
?	2	0	2	2	1	1	2	1	1	0		

O .. Larger appearance of dull coin.

= .. Equality.

N .. Larger appearance of bright coin.

? .. Doubtful.

It is interesting to note that the effect in question is shown in the most marked manner with the silver threepenny bit, which is the smallest coin used ; the farthing comes next. Further increase in size seems to contradict Loewenstein's observation : indeed, with the penny piece the opposite effect is predominant. Hartridge's result that "with perpendicular lighting the dull one appeared to be either the same size or slightly smaller than the bright one. . . ." is consistently contradicted in so far as oblique illumination was more effective in confirming his observation. It is to be remembered, however, that the marginal error may be too large to make any definite statements. The failure of the effect with a white background is shown in Table II. It is also seen that black is the most suitable background. The darkness of the violet probably accounts for

TABLE II.

The Effect of the Background.

	Black.	Violet.	Blue.	Green.	Yellow.	Red.	White.
O	16	10	8	5	7	7	1
=	10	19	11	9	11	15	20
N	6	3	11	16	13	9	11
?	0	0	2	2	1	1	0

its efficiency. Table III. demonstrates that the colour of the light does not seem to affect the general results. It is clear that in the compilation of Tables I. -III. summations had to be carried out in every case: thus

TABLE III.  
The Effect of the Filters.

	Red.	Blue.	Green.	Neutral.
O	14	13	14	13
=	23	25	22	25
N	16	18	19	16
?	3	0	1	2
(a) Large Coins.				
O	6	2	2	0
=	13	16	13	18
N	9	10	12	9
?	0	0	1	1
(b) Small Coins.				
O	8	11	12	13
=	10	9	9	7
N	7	8	7	7
?	3	0	0	1

only the sizes of the coins were considered in Table I. and the results treated as if there had been only one filter and one background. While such a procedure is less satisfactory than a statistical one, it was the only way which could be adopted in the present investigation.

The second part of the experiment\*, on the other hand, was suited for observations by a large number of people. Here nickel threepenny bits were the only coins to be dealt with. To test the correctness of Prof. Hartridge's conclusion, namely that the effect "is either due to an actual difference in size of the two coins or to the way that light is reflected from them," four nickel threepenny bits coined in 1942 were

---

\* Weil, 'Nature,' clvi., p. 636 (1945).

selected. Two of them had their edges and heads filed away\* so that a flat surface was obtained on both. One coin of each pair was carefully polished, attention being paid to the outer regions as before. The other filed coin was brought into contact with plasticene, which immediately tarnished its surface. This was more satisfactory than oxidation by heating, since the risk of deforming the coin was eliminated. The remaining coin was dull and needed no attention.

The filed and the normal coins were placed in pairs on the same neutral background and illuminated by the same parallel beam as before, the same filters being used. Nineteen persons were told that the respective pairs consisted of reputedly equal coins, and everyone put down his results for eight observations: four each for perpendicular and oblique illumination. They were not permitted to discuss their observations while the experiment lasted.

It can be seen from Table IV. that, in the case of oblique lighting, both bright coins appeared to be the larger in all cases except for the filed coin

*The Effect of Filters (for Nickel Threepenny Bits).*

TABLE IV.

Oblique Illumination.

	Red.		Green.		Blue.		Neutral.		Sum.		
									Normal	Filed.	Total.
O	3	4	5	4	6	5	5	8	19	21	40
=	4	4	5	3	3	3	5	7	17	17	34
N	12	11	9	12	10	11	9	4	40	38	78

(The first row in each column refers to the ordinary coins, the second to the filed ones.)

in neutral light. This general result is to be expected on account of retinal irradiation. But the observations for perpendicular incidence are much less convincing: in the case of the filed coins the dull appeared to be the larger in all except blue light, and there the difference is too small to defy confirmation. Actually it might be expected that this exception was spurious, since it was seen in the former experiment that the colour of the filter had no effect on the phenomenon under discussion. Moreover, in the instances where the dull coin appeared to be the larger, the filed one was very much stronger in exhibiting the effect than the normal threepenny bit. This is rather remarkable since perpendicular illumination might be expected sensibly to diminish the appearance of unevenness of the normal coin, as all shadows are greatly reduced.

---

\* Permission to mutilate the coins was obtained from the Deputy Master and Comptroller of the Royal Mint.

The summations on the right of Tables IV. and V. show that, while there is little reason to make any difference between the filed and the normal coins in the case of oblique illumination, the results for perpendicular lighting make it very doubtful if an explanation can be accepted

TABLE V.  
Perpendicular Illumination.

	Red.		Green.		Blue.		Neutral.		Sum.		
									Normal	Filed.	Total.
O	1	9	2	10	0	5	1	8	4	32	36
=	7	7	9	5	2	8	6	8	24	28	52
N	11	3	8	4	17	6	12	3	48	16	64

TABLE VI.  
The Summed Effect of the Filters.

	Red.	Green.	Blue.	Neutral.
O	17	21	16	22
=	22	22	16	26
N	37	33	44	28

if it is based on the manner in which the edges of the coin enter into the problem. Therefore the conclusions of Prof. Hartridge and Mr. Phelps appear to be contradicted, and Prof. Loewenstein's suggestion that there may be a psychological basis for the phenomenon must receive renewed consideration.

The writer would like to thank Miss M. Oliver and Mr. P. Crosby for the patient assistance which they have accorded him throughout the experiment.

---

LXVI. *On a Derivation of the Equations of Equilibrium of a Thin Plate.*

By J. G. OLDROYD, M.A.\*

Courtaulds Ltd., Research Laboratory, Maidenhead, Berks.

[Received June 3, 1946.]

A CONCISE method of deriving the equations of equilibrium of a thin plate, using the principle of virtual work and tensor notation, has been

---

\* Communicated by the Author.



given by Jeffreys<sup>(1)</sup>. But the actual formula obtained for the normal force at the boundary of the plate (equation P(17)\*) does not appear to be consistent with Rayleigh's boundary condition<sup>(2)</sup>

$$\frac{\partial}{\partial n}(\nabla_1^2 w) + (1 - \sigma) \frac{\partial}{\partial s} \left( \frac{\partial^2 w}{\partial n \partial s'} \right) = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

for a free edge, nor with the equation

$$\int \mathbf{N} d\mathbf{s} + \iint \mathbf{Z} d\mathbf{S} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which is required by the equilibrium of the plate as a whole. The following modification is an attempt to avoid the difficulties which inevitably arise when two non-commuting differential operators  $\partial/\partial n$  and  $\partial/\partial s$  are used. The notation below is that of Jeffreys's paper except that, following Rayleigh, cartesian co-ordinates  $n, s'$  are retained in the formulæ,  $n$  being measured along the outward normal and  $s'$  along the tangent to the boundary curve at any particular point of the boundary under consideration. Then the operators  $\partial/\partial n$  and  $\partial/\partial s'$  commute, and  $\partial/\partial s$  is reserved to indicate differentiation along the boundary curve so that, for example,  $\frac{\partial}{\partial n} \left( \frac{\partial w}{\partial s} \right)$  is not defined. There is no ambiguity in using  $n$  and  $s$  as suffixes, and after differentiation  $\partial/\partial s$  and  $\partial/\partial s'$  are interchangeable.

In equation P(8), we can put the scalar quantity

$$l_k \frac{\partial M_{ik}}{\partial x_i} = D l_k \frac{\partial}{\partial x_k} (\nabla_1^2 w) + D \frac{\partial}{\partial n} (\nabla_1^2 w),$$

and the equation can be rewritten (*cf.* equation P(9))

$$\delta V = \int M_{nn} \frac{\partial w'}{\partial n} ds + \int \left\{ M_{ns} \frac{\partial w'}{\partial s} - D \frac{\partial}{\partial n} (\nabla_1^2 w) w' \right\} ds + \iint D \nabla_1^4 w \cdot w' dS. \quad (3)$$

Hence, for equilibrium, after an integration of one term by parts,

$$\begin{aligned} & \int M \frac{\partial w'}{\partial n} ds + \int \mathbf{N} w' d\mathbf{s} + \iint \mathbf{Z} w' d\mathbf{S} \\ & \equiv \int M_{nn} \frac{\partial w'}{\partial n} ds - \int \left\{ \frac{\partial M_{ns}}{\partial s} + D \frac{\partial}{\partial n} \nabla_1^2 w \right\} w' ds + \iint D \nabla_1^4 w \cdot w' dS. \end{aligned} \quad (4)$$

Equations P(10) and P(11), namely

$$\mathbf{Z} = D \nabla_1^4 w \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and

$$\mathbf{M} = M_{nn} = D \left( \frac{\partial^2 w}{\partial n^2} + \sigma \frac{\partial^2 w}{\partial s'^2} \right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

---

\* References to Dr. Jeffreys's paper will be abbreviated in this way.



and similar ones for derivatives of  $\beta$ . Hence

$$\begin{aligned}\frac{\partial}{\partial x} &= \frac{1}{h} \frac{\partial}{\partial \alpha}, \quad \frac{\partial}{\partial y} = \frac{1}{k} \frac{\partial}{\partial \beta}, \\ \frac{\partial^2}{\partial x^2} &= \frac{1}{h^2} \frac{\partial^2}{\partial \alpha^2} - \frac{1}{h^3} \frac{\partial h}{\partial \alpha} \frac{\partial}{\partial \alpha} + \frac{1}{hk^2} \frac{\partial h}{\partial \beta} \frac{\partial}{\partial \beta}, \\ \frac{\partial^2}{\partial x \partial y} &= \frac{1}{hk} \left( \frac{\partial^2}{\partial \alpha \partial \beta} - \frac{1}{h} \frac{\partial h}{\partial \beta} \frac{\partial}{\partial \alpha} - \frac{1}{k} \frac{\partial k}{\partial \alpha} \frac{\partial}{\partial \beta} \right), \\ \frac{\partial^2}{\partial y^2} &= \frac{1}{k^2} \frac{\partial^2}{\partial \beta^2} + \frac{1}{h^2 k} \frac{\partial k}{\partial \alpha} \frac{\partial}{\partial \alpha} - \frac{1}{k^3} \frac{\partial k}{\partial \beta} \frac{\partial}{\partial \beta}.\end{aligned}$$

The operators on the right of these equations do not commute but can be used to differentiate such absolute combinations as  $\nabla_1^2 w$ , which can be defined without reference to a co-ordinate system, without any ambiguity.

The co-ordinates  $x, y$  become  $n, s'$  at the boundary  $\alpha = \alpha_1$ , and we have finally

$$Z = D \nabla_1^4 w, \quad \dots \quad (9)$$

$$M = D \left\{ \left( \frac{1}{h^2} \frac{\partial^2 w}{\partial \alpha^2} - \frac{1}{h^3} \frac{\partial h}{\partial \alpha} \frac{\partial w}{\partial \alpha} + \frac{1}{hk^2} \frac{\partial h}{\partial \beta} \frac{\partial w}{\partial \beta} \right) + \sigma \left( \frac{1}{k^2} \frac{\partial^2 w}{\partial \beta^2} + \frac{1}{h^2 k} \frac{\partial k}{\partial \alpha} \frac{\partial w}{\partial \alpha} - \frac{1}{k^3} \frac{\partial k}{\partial \beta} \frac{\partial w}{\partial \beta} \right) \right\}_{\alpha=\alpha_1}, \quad (10)$$

$$-N = D \left\{ \frac{1}{h} \frac{\partial}{\partial \alpha} (\nabla_1^2 w) + (1-\sigma) \frac{1}{k} \frac{\partial}{\partial \beta} \left( \frac{1}{hk} \frac{\partial^2 w}{\partial \alpha \partial \beta} - \frac{1}{h^2 k} \frac{\partial h}{\partial \beta} \frac{\partial w}{\partial \alpha} - \frac{1}{hk^2} \frac{\partial k}{\partial \alpha} \frac{\partial w}{\partial \beta} \right) \right\}_{\alpha=\alpha_1}, \quad (11)$$

$$\text{where } \nabla_1^2 = \frac{1}{hk} \left\{ \frac{\partial}{\partial \alpha} \left( \frac{k}{h} \frac{\partial}{\partial \alpha} \right) + \frac{\partial}{\partial \beta} \left( \frac{h}{k} \frac{\partial}{\partial \beta} \right) \right\}.$$

(The sign of  $N$  is correct only if  $\alpha$  increases along the outward normal.)

If, as often happens, suitable co-ordinates  $\alpha, \beta$  can be obtained from cartesian co-ordinates by a transformation of the type  $x+iy=f(\alpha+i\beta)$ , the equations simplify a little in virtue of the fact that  $h=k$ .

### References.

- (1) Jeffreys, H., 'A Derivation of the Equations of Equilibrium of a Thin Plate,' *Phil. Mag.* (7), xxxii. p. 365 (1941).
- (2) Rayleigh, 'The Theory of Sound,' i. p. 357, equation (5) with slightly different notation.

## LXVII. On Jeans' Dynamical Theory of Gases.

To the Editor of the *Philosophical Magazine*.

GENTLEMEN,—

WITH reference to the statement in 'Nature,' of December 28, 1946, p. 951, that in checking numerical calculations in the fourth edition, 1925, of Sir James Jeans' "Dynamical Theory of Gases" I differed

SER. 7, VOL. 37, NO. 272—SEPT. 1946.

3 B

substantially from his results in a few cases, it seems desirable to mention the following instances :—

In the first two paragraphs of § 449, p. 342, I reckon that the entries  $3 \times 10^{11}$ , 10 metres, “by about  $10^{-6}$ ,” 10,000 kilometres, and 300,000 molecules, should be replaced respectively by  $4.6 \times 10^{11}$ , 6.4 metres, “to about  $2.1 \times 10^{-7}$  of the value at 800 kilometres,” 30,000 kilometres, and 100,000 molecules. And in the last paragraph of § 440, p. 335, my calculations give 0.1, six, and a “rate of about one per cent. in every 600 metres”, instead of .01, five, and a “rate of about one per cent. in every five kilometres.” And there are other discrepancies, both in this Ch. XV. and elsewhere.

In § 407, p. 304, the value for elastic spheres of

$$\frac{\vartheta}{\sigma T} = 2 \left( \frac{R}{e} \right)^2 \frac{1}{J}$$

is given as “ $1.777 \times 10^8$  in electromagnetic units,”  $\vartheta$  being the coefficient of conduction of heat for any substance, and  $\sigma$  the coefficient of electric conductivity for the same substance. The result should be in heat units, and the correct value, based on the figures for  $e$ ,  $J$  and  $R$  given on pp. 8, 11 and 119, is 3.5503. On p. 196 of Jeans’ ‘Introduction to the Kinetic Theory of Gases,’ 1940, the figure is corrected to 3.538, based on later determinations of  $e$  and  $R$  (3.542 by my calculation, or  $1.482 \times 10^8$  in C.G.S. work units).

On p. 404 of the earlier book a quantum-mechanics formula (981) is given for  $\vartheta/\sigma T$ , and on p. 405 calculations by Hertzfeld for five different temperatures ( $T$ ) are quoted. These calculations are said to be based on the formula (981), and are in close agreement with the observed values. But Hertzfeld’s paper (*Ann. d. Phys.* xli., p. 27 (1913)) shows that although, with his arbitrary value of 57.77 for the quantity  $h\nu_0/R$  which appears in the formula, his calculations ratify the form of the formula, the constant multiplier in (981), viz.  $3R^2/e^2$ , which is  $2.228 \times 10^8$  in C.G.S. work units, has been empirically adjusted by Hertzfeld to about  $2.564 \times 10^8$ , so as to fit the observed values. His results cannot, therefore, be said to afford a close confirmation of formula (981). At the five temperatures of  $100^\circ \text{C.}$ ,  $18^\circ \text{C.}$ ,  $0^\circ \text{C.}$ ,  $-50^\circ \text{C.}$ ,  $-170^\circ \text{C.}$ , the observed values for  $(\vartheta/\sigma T) \times 10^{-8}$  are respectively 2.35, 2.32, 2.30, 2.26, 1.85, and Hertzfeld’s figures are 2.363, 2.313, 2.296, 2.236, 1.865; but formula (981) gives, with  $h\nu_0/R=57.77$ , 2.056, 2.008, 1.993, 1.941, 1.619, viz. some 12 to 14 per cent. less than the observed values.

Yours faithfully,

W. F. SEDGWICK.

---

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



LXVIII. *Theory of the High Pressure Mercury Vapour Discharge* (B).

By V. J. FRANCIS, B.Sc., F.Inst.P., A.M.I.E.E.

(Communication from the Staff of the Research Laboratories of  
The General Electric Company Limited, Wembley, England.)\*

[Received April 8, 1946.]

*List of Symbols.*

$r$	Distance from axis of discharge column in cm.
$T$	Temperature in °K. at $r$ .
$\sigma_0$	See equation (1.1).
$E$	Potential gradient along arc in c.g.s. e.m.u.
$P$	Pressure of mercury vapour in atmospheres.
$\lambda(T)$	See equation (1.3).
$\phi(T)$	See equation (1.4).
$\alpha$	A constant, see A, (2.7).
$\lambda_0$	Electron mean free path in cm. reduced to N.T.P.
$K_1$	See equation (1.5).
$A_3$	A constant, see equation (1.2).
$r_1$	See equation (1.6).
$r_0$	Radius in cm. of discharge tube.
$T_w$	Wall temperature of discharge tube, taken to be always 1000° K.
$T_0$	Temperature in °K. along axis of discharge column.
$\beta$	See equation (1.7).
$T_1$	See equations (1.8), (1.9) and fig. 2 of A.
$\mu_0$	See equation (1.7) and fig. 4 of A.
$I$	Arc current in c.g.s. e.m.u.
$M$	Mass of mercury per cm. of discharge column in grams.
$S(x)$	See equation (1.11).
$g(x)$	See equation (1.12).
$L$	The Radiation Efficiency (see section 6).
$B(x)$	The Radiation Brightness (see section 10).

## SUMMARY.

The theory proposed in an earlier paper<sup>(1)</sup> is further developed. An improved method of finding the values of the constants is used and curves are calculated relating the important discharge parameters over a wide range of operating conditions.

1. *Introduction.*

IN an earlier paper<sup>(1)</sup> (hereinafter called A) the energy balance equation for the high-pressure mercury vapour discharge was set up and solved. In setting up the equation two major assumptions were made :—(1) that

\* Communicated by Sir Clifford Paterson, F.R.S.

the convection and conduction losses could be represented by a coefficient proportional to the temperature  $T$ ,

$$\sigma(T) = \sigma_0 \cdot T, \quad \dots \quad (1.1)$$

and (2) that the radiation losses could be accounted for by the assumption of a single mean level. The value of 8 volts (previously used by Elenbaas<sup>(2)</sup>) was taken for the energy of this fictitious level. The energy balance equation was thus obtained in the form

$$\frac{1}{r} \frac{d}{dr} \left\{ r \frac{dT^2}{dr} \right\} = -K_1 E^2 P^{-\frac{1}{2}} \left[ \lambda(T) - \frac{A_3 P^{3/2}}{E^2} \phi(T) \right], \quad \dots \quad (1.2)$$

$$\text{where} \quad \lambda(T) = T^{\frac{1}{2}} 10^{-\frac{2.63 \times 10^4}{T}}, \quad \dots \quad (1.3)$$

$$\phi(T) = \frac{1}{T} 10^{-\frac{4.02 \times 10^4}{T}}, \quad \dots \quad (1.4)$$

$$K_1 = 2.18 \times 10^{-2} \frac{\alpha \lambda_0}{\sigma_0}, \quad \dots \quad (1.5)$$

The constants  $A_3$ ,  $\alpha$ ,  $\lambda_0$ ,  $\sigma_0$  are not dependent upon  $E$ ,  $T$  or  $P$ , but do, of course, involve the properties of mercury vapour.

In solving the equation (1.2), one major assumption was made, namely, that in order to obtain the temperature distribution  $T(r)$ , throughout the arc we may assume that all the energy is evolved within a certain radius  $r_1$ , and that over the area it is uniformly distributed.

The solution of the energy balance equation was then obtained in the form

$$\left. \begin{aligned} T^2 &= T_0^2 - \beta/2 \left( \frac{r}{r_1} \right)^2 & (r < r_1), \\ T^2 &= T_w^2 - \beta \ln \left( \frac{r}{r_0} \right) & (r > r_1) \end{aligned} \right\}, \quad \dots \quad (1.6)$$

$$\text{where} \quad \beta = \frac{K_1 E^2 \mu_0 r_1^2}{2 P^{\frac{1}{2}}}, \quad \dots \quad (1.7)$$

and boundary conditions give also the relations

$$T_1^2 = T_0^2 - \beta/2, \quad \dots \quad (1.8)$$

$$T_1^2 = T_w^2 - \beta \ln \left( \frac{r_1}{r_0} \right), \quad \dots \quad (1.9)$$

where  $T_1$  is given as a function of  $A_3 P^{3/2}/E^2$ , in fig. 2 of A;  $\mu_0$  is given as a function of  $\beta$  and  $A_3 P^{3/2}/E^2$ , in fig. 4 of A. If  $E$  and  $P$  are given, these relations are sufficient to enable us from equation (1.6) to obtain

$$T = T(E, P, r_0, r). \quad \dots \quad (1.10)$$

As well as  $E$  and  $P$  there are two other variables of importance,  $I$  the total current and  $M$  the mass of mercury per unit length of discharge column. These are measured in c.g.s. e.m.u. and in grams respectively. Since the conditions in the discharge are fully specified when  $E$  and

P are known, there must be two relations between E, P, I and M. It was shown in A that these are

$$M = \frac{15 \cdot 37^2 P}{T_w} S\left(\frac{\beta}{2T_w^2}\right), \quad \dots \quad (1.11)$$

and 
$$I = 0 \cdot 137 \left(\frac{\alpha \lambda_0}{K_1}\right) \frac{1}{E \mu_0} g(T_0^2), \quad \dots \quad (1.12)$$

where  $S(x)$  and  $g(x)$  are given in figs. 6 and 7 of A, respectively.

In equations (1.6) to (1.12), which represent the behaviour of the arc, occur three constants,  $K_1$ ,  $A_3$  and  $\frac{\alpha \lambda_0}{K_1}$ . In paper A an attempt was made to find the values of these constants, not by calculation from the atomic constants but by finding values of them from actual measurable properties of the high-pressure mercury vapour arc itself. A rather involved method of calculation was there used and the data on which the calculations were based consisted of only a few point readings. It was pointed out that this procedure was not likely to give the best values of the constants but that the values obtained were good enough to show whether qualitative agreement existed. It is one of the objects of the present paper to obtain better values of these constants, for further work has shown that closer agreement exists between theory and experiment than was expected.

The method used is to solve the equations characterizing the discharge for one required variable in terms of two other variables. In this way families of curves are obtained (which involve three constants as parameters) from which the general behaviour of the discharge can be seen. Comparison of these calculated characteristics with measured characteristics allows estimates of the values of the constants to be obtained, and these constants can then be used to examine quantitatively the agreement between calculation and experiment. It can be shown in this way without calculating any constants at all that the theory explains many of the general properties of the discharge in a satisfactory manner.

## 2. The Relation between any Three of the Discharge Parameters.

Since only two of E, I, M, P and  $T_0$  are independent there will be one equation connecting any three of them. This dependence will, of course, involve  $r_0$  and the constants which ultimately we have to evaluate. It is these equations which characterize the properties of the discharge and which we seek to find. In order to obtain the required relation between any three of the quantities  $T_0$ ,  $r_1$ ,  $\beta$ ,  $\mu_0$ ,  $T_1$ , E, I, M, P, S and  $g$  we must eliminate all but these three quantities from equations (1.7), (1.8), (1.9), (1.11), (1.12) of the present paper and the relations given in figs. 2, 4, 6 and 7 of paper A. The procedure adopted here is illustrated with reference to Table I. For a fixed value of  $A_3 P^{3/2} / E^2$ , in this case  $4 \times 10^8$ , a series of values of  $\beta / 2T_w^2$  is taken in column 1. The reason for

TABLE I.

$$A_3 P^{3/2}/E^2 = 4 \times 10^8.$$

1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.
$\beta/2T_w^2$	$\mu_0$ $\times 10^2$	$\frac{r_0}{\ln r_1}$	$K_1 E^{2/3} P_1^{2/3}$ $\times 10^{-9}$	$r_1^2/r_0^2$	$A_3 K_1 (Pr_0^2)$ $\times 10^{-18}$	$S\left(\frac{\beta}{2T_w^2}\right)$	$M/Pr_0^2$ $\times 10^3$	$A_3 K_1 M$ $\times 10^{-16}$	$T_0^2$ $\times 10^{-6}$	$T_0$	$\log g(T_0^2)$	$g(T_0^2)$ $\times 10^{-4}$	$EI\left(\frac{K_1}{\alpha\lambda_0}\right)$ $\times 10^{-6}$	$A_3 K_1 r_0^{3/2} E$ $\times 10^{-9}$	L
10.0	.35	1.025	11.40	.127	35.8	.260	3.98	14.20	31.5	5610	4.78	6.03	2.36	23.2	.420
12.5	.46	.820	10.90	.193	21.3	.251	3.85	8.20	34.0	5830	5.00	10.00	2.98	15.7	.425
15.0	.60	.682	10.00	.250	15.7	.244	3.74	5.87	36.5	6040	5.20	15.80	3.61	12.5	.430
17.5	.76	.584	9.21	.312	11.8	.237	3.62	4.15	39.0	6250	5.38	24.00	4.33	10.1	.445
20.0	.94	.512	8.51	.359	9.5	.231	3.54	3.36	41.5	6440	5.57	37.20	5.43	8.53	.494
25.0	1.30	.408	7.70	.440	6.98	.218	3.34	2.33	46.5	6820	5.90	79.4	8.36	6.80	.590
30.0	1.56	.342	7.70	.503	6.13	.206	3.16	1.94	51.5	7180	6.17	148	13.00	6.15	.684
35.0	1.58	.292	8.86	.560	6.34	.197	3.02	1.91	56.5	7520	6.42	263	22.80	6.25	.790
40.0	1.50	.256	10.70	.601	7.13	.188	2.88	2.06	61.5	7840	6.62	417	38.10	6.90	.856
45.0	1.34	.227	13.40	.635	8.45	.180	2.76	2.33	66.5	8150	6.81	646	66.20	7.85	.906
50.0	1.19	.205	16.80	.661	10.2	.171	2.62	2.68	71.5	8460	6.96	912	105.00	9.00	.935
55.0	1.08	.185	20.40	.687	11.9	.163	2.50	2.98	76.5	8750	7.06	1150	146.00	10.10	.948
60.0	1.00	.170	24.00	.706	13.6	.155	2.38	3.24	81.5	9030	7.11	1290	177.00	11.20	.954



choosing this particular value of  $A_3P^{3/2}/E^2$  is that this is one of the values for which  $\mu_0$  is plotted in fig. 4 of A.

The values of  $\mu_0$  in column 2 are then obtained from fig. 4 of A. Calculations based on values of  $\beta/2T_w^2$  less than 10 are not to be relied upon, depending as they do on the commencement of the  $\mu_0$  curves, where a small absolute error in  $\mu_0$  is likely to be a large percentage. This will result later in the fact that the calculated characteristics cannot be continued to  $EI \rightarrow 0$ . Fig. 2 of A gives the value of  $T_1$  corresponding to  $A_3P^{3/2}/E^2 = 4 \times 10^8$  as 4630. Using this value and the figures in column 1, we get from equation (1.9) the value for  $\ln(r_0/r_1)$  given in column 3. Column 4 follows from columns 1 and 2 and equation (1.7); and column 5 from column 3. Now using columns 4 and 5 and the relation  $A_3P^{3/2}/E^2 = 4 \times 10^8$ , we obtain column 6. Column 7 is obtained from fig. 6 of A, and column 8 from column 7 and equation (1.11). Columns 6 and 8 then give column 9. We thus have two of our important variables separated:  $P$  in column 6 and  $M$  in column 9. In order to separate the other two variables  $E$  and  $I$ , which we have in effect in columns 14 and 15, we start with equation (1.8), from which we obtain columns 10 and 11, and then fig. 7 of A and equation (1.12) lead to the results in columns 12, 13 and 14. Finally, from column 6 and the known value of  $A_3P^{3/2}/E^2$ , we arrive at column 15. The results in column 16 will be discussed later.

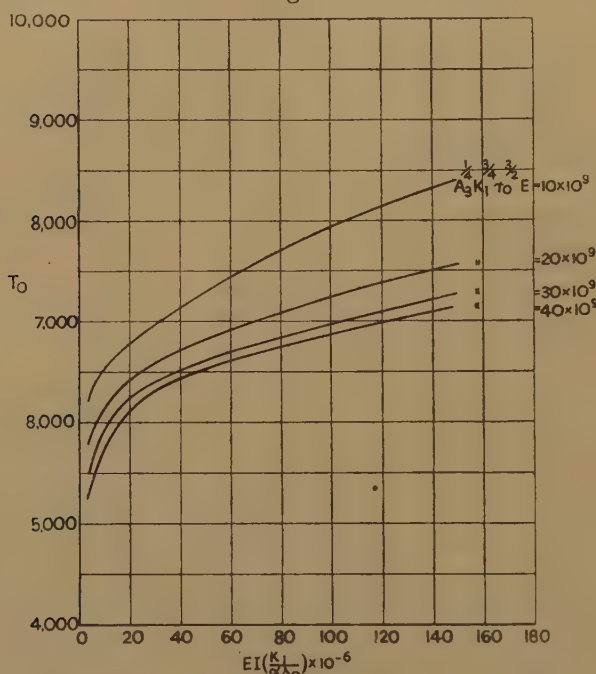
Similar tables were calculated for values of  $A_3P^{3/2}/E^2 \times 10^{-8}$  of 2, 6, 8 and 10. For this purpose it was necessary to extend somewhat the curves given in A; in particular it was necessary to calculate a  $\mu_0$  curve for  $A_3P^{3/2}/E^2 = 8 \times 10^8$ . To save space these other tables are not reproduced here as they serve only the purpose of allowing the results produced below to be calculated. This set of tables gives, in effect, any quantity at the head of a column as a function of the two variables  $A_3P^{3/2}/E^2$  and  $\beta/2T_w^2$ . We may therefore find the relation between any three of the variables at the head of the columns by eliminating  $A_3P^{3/2}/E^2$  and  $\beta/2T_w^2$ . Since the functions involved cannot easily be expressed algebraically, the necessary interpolation must be done either graphically or numerically, the accuracy obtained by graphical methods is probably all that is justified unless numerical tables are used for the  $\mu_0$  and  $g$  functions. In the calculations actually carried out here the values of  $\mu_0$  and  $g$  were obtained from the curves in A, and therefore no great numerical accuracy can be expected. The intervals both of  $\beta/2T_w^2$  and  $A_3P^{3/2}/E^2$  proved to be larger than desirable even for graphical interpolation, and a certain amount of smoothing was necessary; this was particularly the case for the  $M, E, EI$  relation given in section 5, where the somewhat complex nature of the resulting functions really requires that much smaller intervals should be taken. For the present purpose it was not thought worth while to increase considerably the amount of computation necessary in order to produce an accuracy that would have questionable physical significance. There is little doubt that the functions calculated here are accurate to within as small limits as are necessary.

The actual elimination can be carried out in several ways, all of which are, of course, fundamentally the same. The actual procedure adopted is described briefly in the next section, where  $T_0$  is obtained as a function of the watts per centimetre of the discharge ( $EI \times 10^{-7}$ ) for various constant values of  $E$ .

### 3. The Relation between $T_0$ , $EI$ and $E$ .

To obtain this relation from Table I. and the similar tables,  $T_0$  (from column 11) was plotted against  $A_3 P^{3/2}/E^2$  for each value of  $\beta/2T_w^2$ . A similar family of curves was obtained from  $EI(K_1/\alpha\lambda_0)$  in column 14 against  $A_3 P^{3/2}/E^2$ , and also for  $A_3^{\frac{1}{2}} K_1^{\frac{1}{2}} r_0^{3/2} E$  in column 15. It is not necessary to reproduce these curves here as they have no significance

Fig. 1.



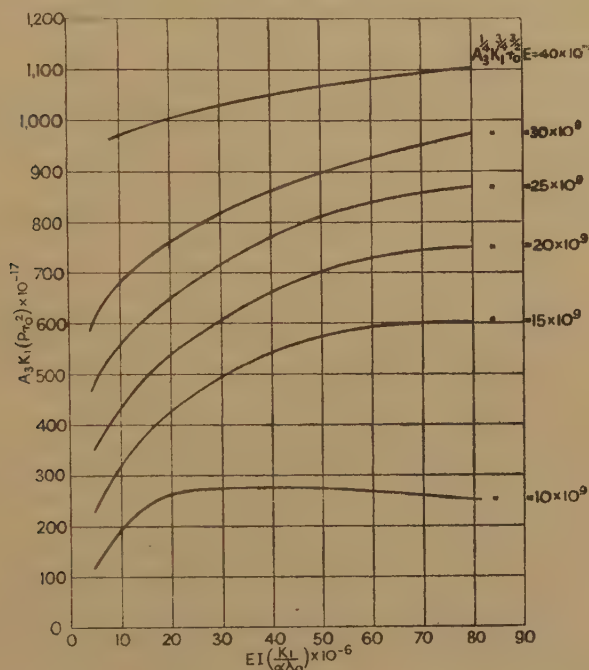
apart from their use as an instrument in the calculations. From the family of curves relating  $A_3^{\frac{1}{2}} K_1^{\frac{1}{2}} r_0^{3/2} E$  and  $A_3 P^{3/2}/E^2$  a constant value of the former was taken, and for each value of  $\beta/2T_w^2$  the corresponding value of  $A_3 P^{3/2}/E^2$  was obtained. Then the values of  $T_0$  and  $EI(K_1/\alpha\lambda_0)$  for each pair of values of  $A_3 P^{3/2}/E^2$  and  $\beta/2T_w^2$  were taken from their respective families of curves. In this way the family of curves shown in fig. 1 was obtained relating  $T_0$ ,  $EI$  and  $E$ . The trend of the curves is much what would be expected. The rapid decrease of temperature with increase of arc voltage at constant loading, however, is probably not easily predictable from general considerations. As far as the writer is aware, there are no reliable experimental results which would indicate the trend of  $T_0$  with other parameters, so that the curves in fig. 1 do not, in

themselves, offer any confirmation that the theory bears any relation to the properties of the mercury discharge. It should also be remembered that the device used in section 3 of A for the solution of the energy balance equation will result in a smaller calculated value of  $T_0$  than that actually obtaining. This effect is not likely to be large, however, and the shape of the curves would be affected only to a second order.

#### 4. The Relation between P, EI and E.

The process here was similar to that used in section 3. The results in columns 6, 14 and 15 were used in this way to obtain the curves shown in fig. 2. Again, so far as the writer is aware, there is no experimental

Fig. 2.



evidence with which to compare the general shape of these curves except, of course, the obvious fact that over a wide range of conditions the pressure increases both with  $E$  and with  $EI$ . No very useful appeal can therefore be made to them for confirmation of the theory.

#### 5. The Relation between M, EI and E.

This is one of the most important relations between the discharge parameters, partly because of its practical importance in the design and operation of high-pressure mercury discharge tubes and partly because this is the only relation which is at all easily accessible to experimental confirmation. Unfortunately, over the region in which we are interested, the curves are somewhat more complex in shape than those dealt with

in the last two sections and, as already mentioned, this leads to added difficulty in a rough computation of them. The method used was the same as that in the preceding two sections, columns 6, 14 and 15 being

Fig. 3.

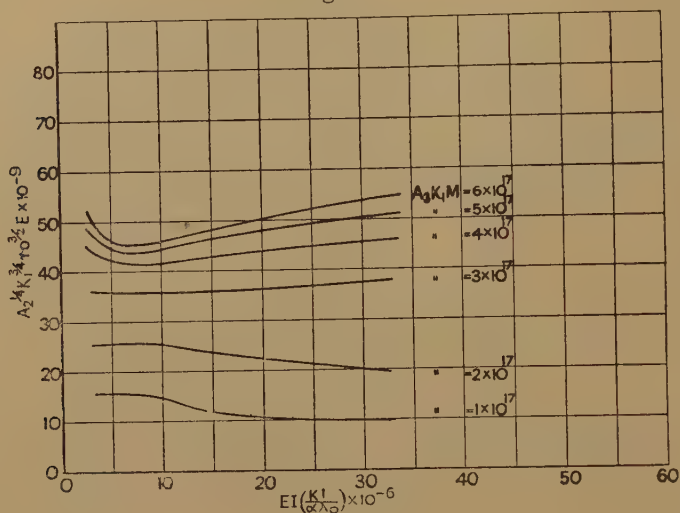
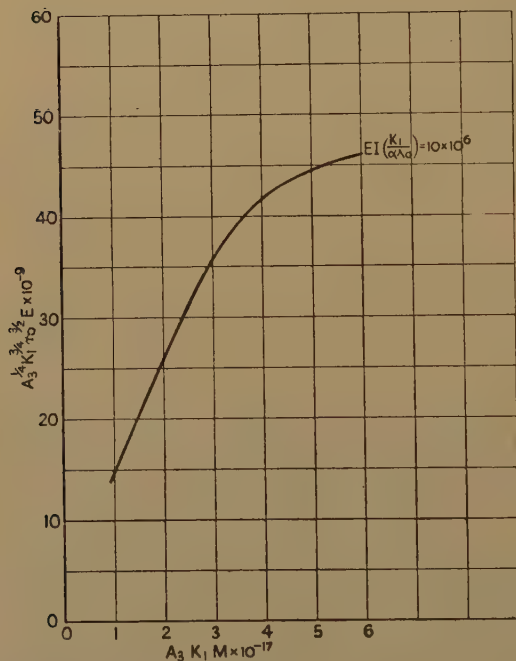


Fig. 4.



used to determine the required family of curves. These are shown in figs. 3 and 4, and there is no doubt that the calculations are sufficiently accurate for our present purpose except at the smaller values of  $EI$ . In



this case, in contrast to the relation dealt with in sections 3 and 4, something is known about the shape of these curves. It is well known that, at least for a certain range of  $M$ , there is a minimum value of  $E$  for some value of  $EI$ ; this is well in evidence for the upper curves in fig. 3. It is also well known that over a very wide range of  $EI$ ,  $E$  varies very little with  $EI$  at constant  $M$ . The fact that for the smaller values of  $M$  there is no pronounced minimum is a little surprising.

In fig. 4  $E$  is given as a function of  $M$  at constant  $EI$  for the value  $EI(K_1/\alpha\lambda_0)=10 \times 10^6$ . The shape of these curves is also well known—an almost linear increase at the lower values of  $M$  and a gradual turn over at the larger values of  $M$ . Both of these properties are clearly evident in fig. 4. These curves therefore provide the first evidence that qualitatively the trend of the functions is given correctly by the theory. This matter will be dealt with in more detail in section 8.

### 6. The Radiation Efficiency.

We define the " radiation efficiency " ( $L$ ) here as

$$L = \frac{\text{Energy Radiated}}{\text{Energy Input}}.$$

It follows at once from (2.3), (2.11), (2.14), (2.17) and (3.3) of A that

$$\begin{aligned} L &= \frac{\int_0^{r_0} 2\pi r \cdot \frac{K\sigma_0}{2} A_3 \cdot P\phi(T) dr}{\int_0^{r_0} 2\pi r \cdot 1.09 \times 10^2 \cdot \alpha\lambda_0 \cdot E^2 P^{-\frac{1}{2}} \lambda(T) dr} \\ &= \frac{A_3 P^{3/2}/E^2 \int_0^{r_0} r\phi(T) dr}{\int_0^{r_0} r\lambda(T) dr} \\ &= 1 - \frac{\int_0^{r_0} r\mu(T) dr}{\int_0^{r_0} r\lambda(T) dr}, \quad \dots \dots \dots (6.1) \end{aligned}$$

and then from (4.3) and (6.6) of A we have

$$L = 1 - \frac{\beta\mu_0}{2g(T_0^2)}, \quad \dots \dots \dots (6.2)$$

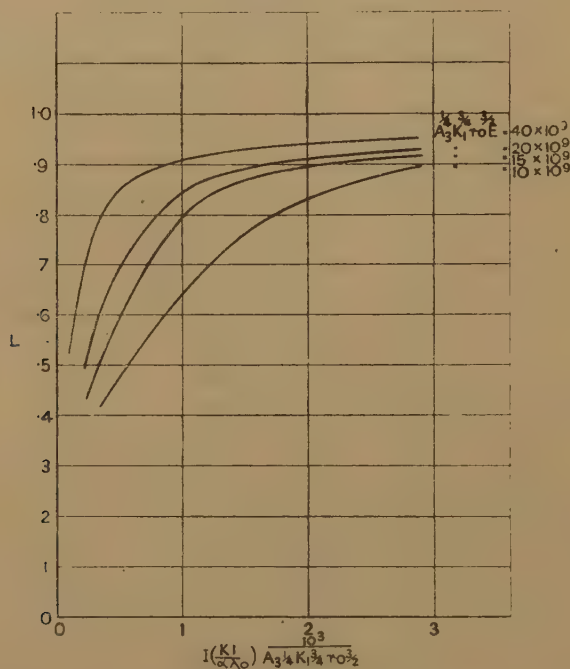
or it may also be expressed as

$$L = 1 - \frac{0.0686\beta}{EI \left( \frac{K_1}{\alpha\lambda_0} \right)}. \quad \dots \dots \dots (6.3)$$

We may therefore obtain  $L$  as a function of  $A_3 P^{3/2}/E^2$  and  $\beta$  from columns 1, 2 and 13 of Table I. and the similar tables for other values of  $A_3 P^{3/2}/E^2$ . The results for  $A_3 P^{3/2}/E^2 = 4 \times 10^8$  are given in column 16 of

Table I. The chief interest is in knowing  $L$  as a function of  $E$  and  $EI$ . This may be obtained in the same way as the other relations were found, and the results are shown in fig. 5. It is convenient here, in order to spread the curves out, to plot  $L$  against  $I$  for constant  $E$ . The general shape of the curves is what would be expected; for all values of  $E$  it appears that  $L=1$  is approached asymptotically. This, of course, is reasonable, for at the value  $L=1$  there would be no energy left to heat the vapour. There are no published data on the total discharge radiation as a function of  $E$  and  $I$ , but the rapid increase in  $L$  would perhaps be

Fig. 5.



expected from the known behaviour of the luminous efficiency in this region.

### 7. Grouping of the Constants and Variables.

Conclusions of some interest can be drawn from the shape and form of the characteristics already calculated. In the first place the constants  $A_3$ ,  $K_1$  occur only in the form of  $A_3^{1/2} K_1^{3/2}$  and  $A_3 K_1$ . It should be noted that the constant  $K_1/\alpha\lambda_0$  may be replaced by a constant involving only  $\sigma_0$  in virtue of equation (1.5). There does not appear to be any obvious physical significance to be attached to the fact that  $A_3$  and  $K_1$  group in the two forms above; it does mean, however, that  $A_3^{1/2} K_1^{3/2}$ ,  $A_3 K_1$  and  $\sigma_0$  are the natural constants, and the attempt made in A to estimate  $A_3$  and  $K_1$  instead of  $A_3^{1/2} K_1^{3/2}$  and  $A_3 K_1$  would be likely to lead to error.

It appears also that the natural manner for the effect of the parameter  $r_0$  to appear is in the form  $Pr_0^2$  and  $Er_0^{3/2}$ , and if these quantities are

observed instead of simply  $P$  and  $E$  the effect of various tube diameters is automatically taken into account. In section 9 an attempt is made to see to what extent this conclusion is borne out in practice. Provisionally, however, we may expect that the important variables are  $M$ ,  $EI$ ,  $Pr_0^2$ ,  $Er_0^{3/2}$  and the effective constants  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$ ,  $A_3K_1$  and  $\sigma_0^{(3)}$ .

### 8. Values of the Constants.

In paper A the attempt was made to estimate the values of the constants  $A_3$  and  $K_1$  from single point readings without first calculating the forms of the characteristics. This is an unsatisfactory process since such constants will be dependent on the experimental accuracy of a few points. The procedure adopted here is much better in that it does enable the values of the constants to be based on a much wider range of data.

Since  $E$ ,  $I$  and  $M$  are the quantities most easily measured, one would hope that all the constants could be obtained by an appeal to characteristics depending on these quantities only. Unfortunately, so far as experimental values at present available are concerned, this is not the case. There are two reasons for this: one is that the comparatively unknown factor of electrode losses makes an accurate estimate of  $E$  impossible, and the other is that, from the way in which the constants occur, it appears that fairly small differences in  $E$  can cause large differences in the resulting estimated values.

It may appear surprising that so little is known about the magnitude of electrode losses; but these losses will depend on many factors including the design of electrode, the tube diameter, the wattage per centimetre of arc and the voltage per centimetre of arc. Whereas with a particular standardized design of electrode the losses could be obtained experimentally for given tube diameter and wattage and voltage loading of the arc with a reasonable amount of work, the amount of experimental work required for an exhaustive examination is very large indeed. Fortunately for arc lengths of several centimetres the errors introduced by our lack of knowledge in this respect are not likely to be large, but it makes deductions from experimental results on short arcs precarious. A further difficulty is that over most of the characteristics of fig. 3,  $E$  is substantially independent of  $EI$ , which makes an accurate estimate of  $K_1/\alpha\lambda_0$  very difficult.

In fig. 6 is shown the  $r_0^{3/2}E-EI$  characteristic for a tube 34 mm. internal diameter ( $r_0=1.7$ ). This was deduced from the volt-ampere characteristic assuming that, at the normal operating condition ( $EI=22.5 \times 10^7$ ), the electrode losses were 40 watts, and making the further reasonable assumption that for other operating conditions the electrode losses were proportional to  $EI$ . For a tube of the same dimensions in fig. 7 (curve  $X_1$ ) is given the measured relation between  $M$  and  $r_0^{3/2}E$ , making the same assumption with regard to electrode losses.

Now, it is known that for all points on curve  $X_1$  in fig. 7 the operating conditions are near the minimum of the corresponding curve representing

the E—EI relations of fig. 3, and a comparison of fig. 6 and fig. 3 shows that, for this value of  $M$  ( $\cdot 0115$ ),  $A_3 K_1 M$  is probably of the order of 5 or

Fig. 6.

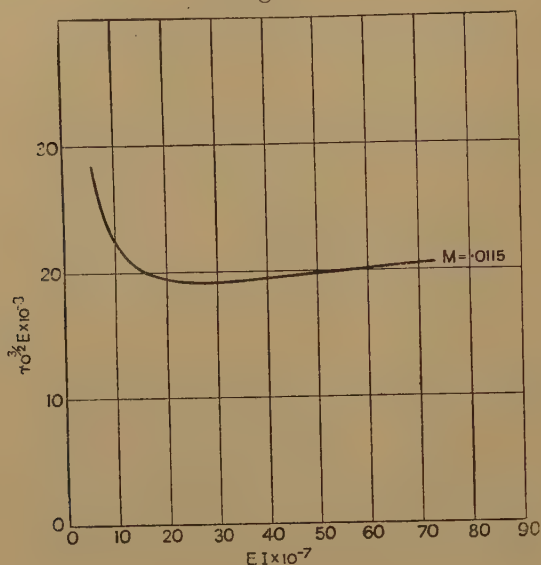
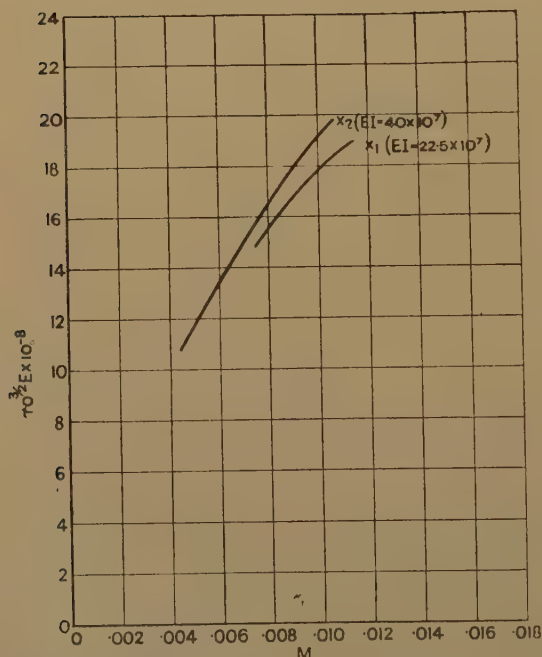


Fig. 7.

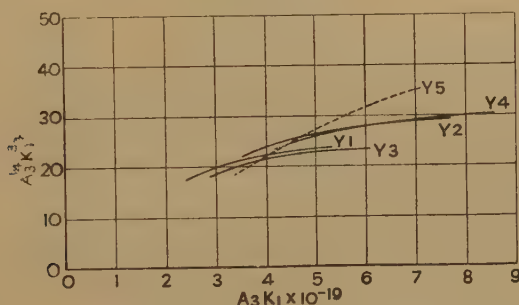


$6 \times 10^{17}$ . We therefore find from the minima of the curves in fig. 3 that, for  $M = 0.0115$  and  $r_0^{3/2} E = 19 \times 10^8$ , if  $A_3 K_1 M = 6 \times 10^{17}$ , then  $A_3^{1/2} K_1^{1/2} r_0^{3/2} E = 45.3 \times 10^9$ ; similarly, if  $A_3 K_1 M = 5 \times 10^{17}$ , then



$A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}r_0^{3/2}E = 43.8 \times 10^9$ . In this way we obtain the relation between  $A_3K_1$  and  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$  shown in fig. 8 (curve  $Y_1$ ). Similarly, if we take a point at the other end of curve  $X_1$  in fig. 7,  $M = .008$ ,  $r_0^{3/2}E = 15.5 \times 10^8$ ; knowing that this operating point is also near the minima of the curves in fig. 3, we may, to the accuracy to which the measured values justifies us in working, obtain in the same way the curve  $Y_2$  in fig. 8. Formally these two curves  $Y_1$  and  $Y_2$  should allow us to obtain the constants  $A_3K_1$  and  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$  by their point of intersection. The curves are so nearly parallel, however, that a relatively small error in either the calculated curves or the measured values, results in a large error in the value of  $A_3K_1$ . This is further illustrated by using characteristics from a tube of very different dimensions and operating conditions. In this case  $r_0 = 0.675$  and (assuming a loss of 40 watts at the electrodes)  $EI = 49 \times 10^7$ . The measured relation between  $M$  and  $r_0^{3/2}E$  is given in fig. 7 (curve  $X_2$ ). Now even a tube of this type, for all the range of  $M$  in fig. 7, operates so near the minimum of the  $E$ — $EI$  curves that, if we assume that the characteristic  $X_2$  in fig. 7 applies to the minima of the

Fig. 8.



curves of fig. 3, any error that we make on this account will be much smaller than other unavoidable inaccuracies. Applying, therefore, the same procedure to two extreme points of the curve  $X_2$  as was used to obtain  $Y_1$  and  $Y_2$  in fig. 8, we obtain the curves  $Y_3$  and  $Y_4$  in fig. 8.

It is clear, therefore, that these curves alone do not provide us with a means of obtaining the constants. There is nothing wrong, of course, in the fact that they do not; in principle there is no reason why they should not all coincide, in which case they would give no indication at all of the values of  $A_3K_1$  and  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$ . From one point of view, in fact, it may be regarded as satisfactory that over such a wide range of operating conditions curves obtained in such a way do so nearly coincide. It does mean, however, that in order to obtain the values we want we must introduce some other parameter. The only other one we have on which any experimental data are available is  $P$ . To use this parameter we have to turn to the characteristics of fig. 2, and this means that we must have the value of  $K_1/\alpha\lambda_0$ . The only way in which we can obtain this is from the position of the minima in fig. 3; and since these minima are very flat, it is not a very satisfactory method. However, a comparison

of figs. 6 and 3 (assuming, as will presently be verified, that the upper curves of fig. 3 are the ones with which to make comparison) that, when  $EI(K_1/\alpha\lambda_0)=7\times 10^6$ ,  $EI=30\times 10^7$ , from which it follows that

$$\frac{\alpha\lambda_0}{K_1}=43 \text{ (approx.)} \quad (8.1)$$

Elenbaas' data<sup>(4)</sup> for his experiment involving a direct measurement of P and M are as follows :—

$$\left. \begin{array}{l} P=0.95 \\ EI=43.8\times 10^7 \\ E=8.0\times 10^8 \\ r_0=1 \\ M=0.00306 \end{array} \right\} \quad (8.2)$$

In paper A, approximate values of some of these quantities were used, but it is thought best here to use the most accurate values available. We have then from (8.2) that

$$\begin{aligned} EI(K_1/\alpha\lambda_0) &= 10.2\times 10^6, \\ r_0^{3/2}E &= 8\times 10^8, \\ Pr_0^2 &= 0.95. \end{aligned}$$

Thus, from each curve in fig. 2 we obtain a pair of values of  $A_3K_1$  and  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$ . These values are plotted in fig. 8 (curve  $Y_5$ ).

It appears then that from fig. 8,  $A_3K_1$  lies between  $4\times 10^{19}$  and  $5\times 10^{19}$ . Comparison of fig. 6 with fig. 3 at the smaller values of EI suggests that the larger value of  $A_3K_1$  would give the better fit, particularly as it is known that there is a pronounced upward bend to the EI—E curves at the lower values of EI, even for M as small as 0.005. We therefore take as the best values of the constants :—

$$\left. \begin{array}{l} A_3K_1=5\times 10^{19} \\ A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}=27 \\ \alpha\lambda_0/K_1=43 \end{array} \right\} \quad (8.3)$$

It is of interest to note that the values of  $A_3$  and  $K_1$  obtained in paper A would give for  $A_3K_1$   $3.15\times 10^{19}$  and for  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}$  22.5.

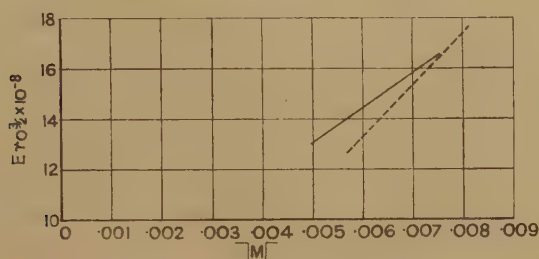
### 9. Further Comparison with Experimental Results.

The results obtained so far—the general form of the characteristics calculated and the fact that from several independent sets of data covering widely different operating conditions it proves possible to obtain approximately the same value of the constants—give a good indication that over a wide range of conditions the agreement is satisfactory from a qualitative point of view. It is, however, worth while to make further examination to see how exact this agreement is.

If, for example, we take the commercial 125 w. quartz lamp with  $r_0^{3/2}E=7.4\times 10^8$ ; it is clear that EI will be approximately  $40\times 10^7$ , so that

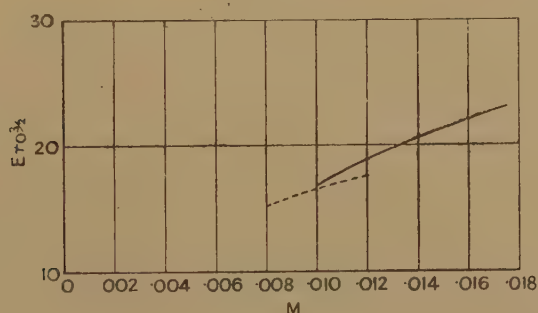
$EI(K_1/\alpha\lambda_0) \times 10^{-6}$  will be nearly 10 and a small error here will be insignificant in reading  $A_3^{\frac{1}{2}}K_1^{\frac{1}{2}}r_0^{3/2}E$ . We have then  $E = 29.6 \times 10^8$ . The measured value of the voltage across the arc was 125, so that agreement exists only if the voltage loss at the electrodes is about 35, which would be also a wattage loss of 35. At first sight this would seem almost certainly too high for this type of tube, but it is difficult to say how much too high. It should be noted that for the 34 mm. diameter tube used in section 8 the watts lost for  $EI=30$  would have been assumed to be  $\frac{30}{22.5}$  or 53.

Fig. 9.



In fig. 9, for a tube with  $r_0=1.18$  operated in the region of  $EI=15 \times 10^7$ , is shown as the full line the  $M-Er_0^{3/2}$  characteristic deduced from the measured arc voltage; a wattage loss at the electrodes equal to  $EI$  was assumed. To obtain the calculated characteristic it is necessary to

Fig. 10.



extrapolate the curves of fig. 3 for the small value of  $EI$  required—a somewhat dangerous proceeding in view of what has already been said about the accuracy of the curves at the lower end. However, when this is done the dotted curve in fig. 9 results. The agreement, in the circumstances, must be considered as satisfactory.

Fig. 10 shows a similar comparison for a tube of very different type operated under quite different conditions. In this case  $r_0=2.15$  and  $EI$  is in the neighbourhood of  $35 \times 10^7$ . The results deduced from the measured values of the arc voltage, assuming a loss at the electrodes of 35 watts, is shown as the full line. The range covered by the curves of

fig. 3 do not allow a comparison over the whole of the curve, but the dotted curve in fig. 10 obtained from the upper three curves of fig. 3 must again be accepted as good agreement.

Elenbaas<sup>(5)</sup> obtained experimentally, over a very wide range of operating conditions,  $E$  as a function of  $P$  and  $I$ . These characteristics are not so useful for our purpose as his experimental result involving  $P$  which was used above, since the values of  $M$  are not known explicitly.

Fig. 11.

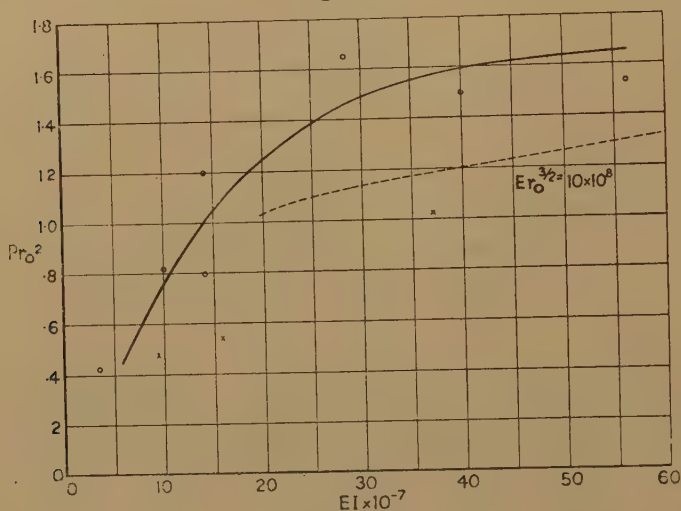
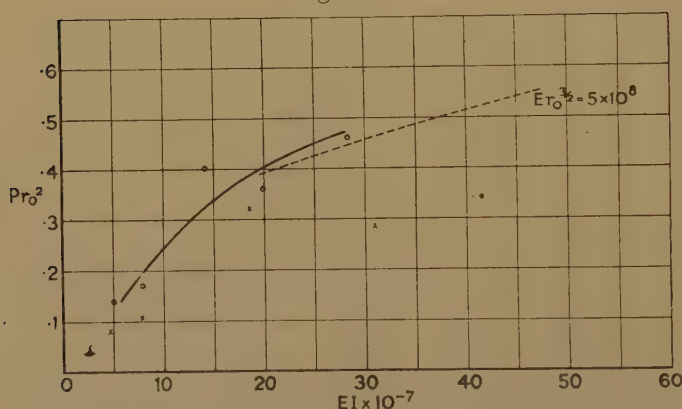


Fig. 12.



However, from figs. 2 to 6 of Elenbaas' paper we are able to deduce the relations between our variables  $Pr_0^2$ ,  $EI$ , and  $Er_0^{3/2}$ . In figs. 11 and 12 of the present paper are shown plotted as circles the points deduced from Elenbaas' results on tube diameters of 40 mm., 20 mm. and 10 mm. for  $Er_0^{3/2} = 10 \times 10^8$  and  $5 \times 10^8$  respectively. The crosses refer to the experimental results on tubes of 4.5 mm. and 2 mm. diameter. The full curve in each case has been drawn through the circles. At first



sight the points—which should, of course, according to the present theory all lie on a single curve—appear to be very badly scattered about any line that can be drawn through them. This is particularly the case if the crosses are considered as well as the circles. It must, however, be borne in mind that the range of operating conditions represented by these points is enormous. For example,  $r_0$  ranges from 0.1 to 2.0 and  $P$  from about 0.02 to 100. Considering the possibility of experimental error and the strain on any simple theory to cover such a range, it would appear that there is good additional evidence here that particular significance attaches to  $Pr_0^2$  and  $Er_0^{3/2}$  as variables. One point, however, is worth noting: the crosses referring to the tubes of smaller diameter are also those referring to the very much higher pressures. It is in these circumstances, when continuous radiation begins to play an important part, that we should expect the line radiation term used in the present calculations to depart from the actual conditions. This is mentioned again in section 11.

So far as support for the values of the constants deduced is concerned, the position can be seen by the dotted curves in figs. 11 and 12, which are taken from fig. 2. The agreement is good in fig. 12 and not so good in fig. 11. It would perhaps be possible to obtain a better fit by modifying the values of the constants already found, but at the present stage it does not appear worth while, to attempt to obtain better agreement in this way, as it appears that the calculations do all that can reasonably be expected with the values already obtained. It still does not seem possible at present to decide how much of the discrepancies existing are due to the crudeness of the assumptions made in developing and solving the equations, and how much is due to experimental error or to assumptions made in deducing the values of the discharge variables from the experimental results. In general, the agreement between measured and calculated values over wide ranges of the variables is better than would be expected. It would be easy to obtain very small discrepancies over a restricted field in which the constants were obtained, but deviations would be expected as conditions departed widely from these. In view of possible experimental errors it does not appear possible to decide how large these discrepancies in fact are.

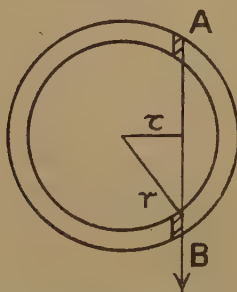
#### 10. *The Radiation Brightness and the Constriction of the Discharge Column.*

It is worth while to consider the distribution of radiation across the discharge column. This would be an extremely useful method of comparing theory with experiment if there existed a close correlation between the radiation efficiency and radiation brightness and the luminous efficiency and the actual brightness. As it is, there is no justification for assuming such a correlation, even from the point of view of line radiation alone, neglecting the effects of continuous radiation. This is clear when one considers that the population of the fictitious level is not the same function of the discharge tube parameters as the

population of the actual levels giving rise to visible radiation. For example, if only one level were responsible for the visible radiation, the other levels giving radiation outside the visible spectrum, and if this level had an energy and transition probabilities quite different from the mean values associated with the fictitious level, as it almost certainly would, then the visible radiation would depend upon the temperature in a manner quite different from the total radiation due to the fictitious level. One could perhaps use the temperature distribution calculated from the theory to estimate the transitions giving rise to visible radiation, and this may be worth doing. Here, however, we merely examine some properties of the total radiation.

To find an expression for the distribution of radiation across the discharge, consider the cross-section of the column in fig. 13, and the

Fig. 13.



radiation emitted in a direction perpendicular to the axis of the column at a distance  $x$  from it. The radiation per unit volume emitted in all directions from the discharge at distance  $r$  from the axis is

$$\frac{K_1 A_3 P \sigma_0}{2} \phi(T). \quad (10.1)$$

Assuming no absorption, the radiation from the two portions of the annulus  $dr$  thick along the direction AB is

$$\frac{K_1 A_3 \sigma_0 P}{4\pi} \phi(T).$$

So that the total emitted radiation per unit surface along AB is

$$B(x) = \frac{A_3 K_1 \sigma_0 P}{4\pi} \int_x^\infty \frac{\phi(T) r dr}{\sqrt{r^2 - x^2}}; \quad (10.2)$$

using (1.6), this becomes

$$B(x) = \frac{A_3 K_1 \sigma_0 P r_1}{8\pi} \cdot \sqrt{\frac{2}{\beta}} I \left( \sqrt{T_0^2 - \frac{\beta}{2r_1^2} x^2} \right), \quad (10.3)$$

where

$$I(Z) = 2 \int_0^Z \frac{\phi(T) T dT}{\sqrt{Z^2 - T^2}} \\ = 2Z \int_0^1 \phi(Z\sqrt{1-y^2}) dy. \quad (10.4)$$

$I(Z)$  was found roughly by graphical methods for various values of  $Z$  and the values of  $\log I(Z)$  are given in fig. 14,

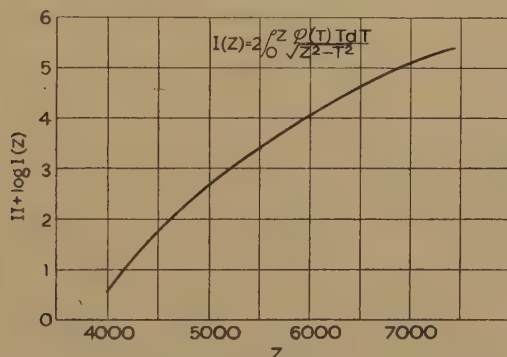
For a given operating condition the only part of  $S(x)$  which depends upon  $x$  is  $I\sqrt{T_0^2 - \frac{\beta}{2r_1^2} \cdot x^2}$ . This expression therefore gives the variation of the radiation brightness (the energy radiated per unit area of surface) across the discharge column. This variation depends only upon  $T_0$  and  $\beta/r_1^2$  and tubes in which  $T_0$  and  $\beta\left(\frac{r_0}{r_1}\right)^2$  are similar will have the same degree of constriction as a function of  $x/r_0$ . In fact, the value of  $x/r_0 = d$  for which the value of  $B(x)$  falls to half its maximum value is given by

$$I(T_0) = 2I\left(\sqrt{T_0^2 - \beta/2\left(\frac{r_0}{r_1}\right)^2 d^2}\right), \quad \dots \quad (10.5)$$

the solution of which gives  $d$  as a function of  $T_0$  and  $\beta/2\left(\frac{r_0}{r_1}\right)^2$ .

It would not be very difficult from (10.5) to obtain a family of curves giving  $d$  as a function of  $EI$  and  $E$ . This is not done here since the results would not throw any further light on the essential properties of

Fig. 14.



the discharge, and it is sufficiently obvious from the form of (10.5) and from the next two paragraphs that the behaviour of  $d$  is much what would be expected.

In order to solve (10.5) for  $d$  in terms of  $E$  and  $EI$  it would be necessary to have  $r_1/r_0$  as a function of these two quantities. This relation has a certain interest on its own account since it gives in effect the extreme limit of the discharge column. It is easily obtained from Table I. and the similar tables for other values of  $A_3 P^{3/2}/E^2$  in the same way as the relations for  $T_0$  and  $P$  obtained in sections 3 and 4. The results are shown in fig. 15. The effect of increasing  $EI$  in increasing  $r_1/r_0$ , and thus spreading the discharge, is expected and is strongly evident. The effect of  $E$  at constant  $EI$  is much less. In this case it would be expected that increase of  $E$  at constant  $EI$ , which (for constant  $r_0$ ) can only be obtained by increasing  $M$ , would be to increase the constriction. If  $r_1/r_0$  is a measure of the observed constriction, which is not necessarily the case, it would seem that there are regions in which increase of  $E$  is accompanied by increase of  $r_1/r_0$ . It is not known definitely whether this is, in fact,

true ; but of course it must be remembered that  $r_1/r_0$  is essentially a measure of the complete diameter of the column, and the observed actual width of column may not always increase or decrease with  $r_1/r_0$ . The matter is worth further examination as it does appear possible that the constriction is a more complicated matter in certain ranges of operation than merely increasing with increase of  $E$  or decrease of  $EI$ .

Fig. 15.

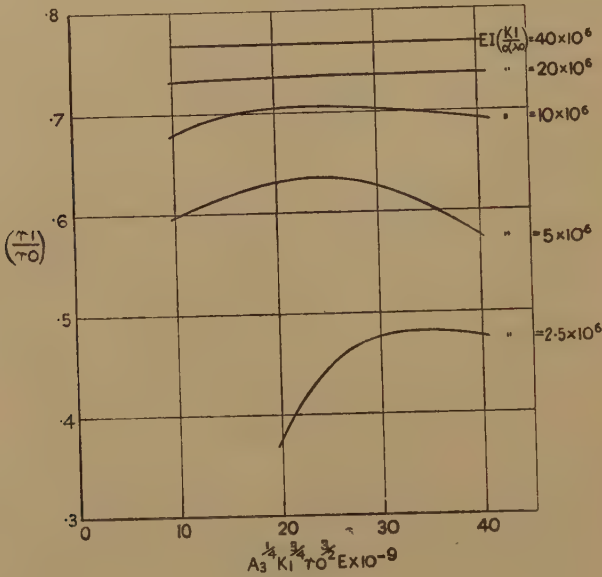
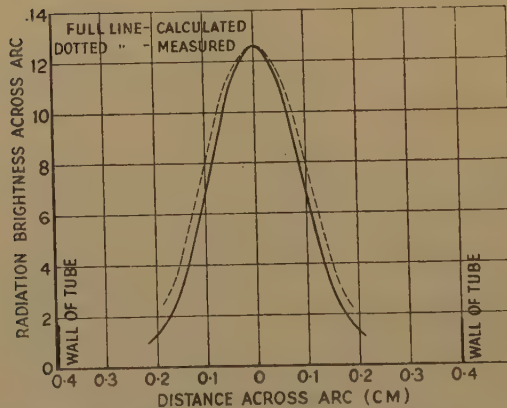


Fig. 16.



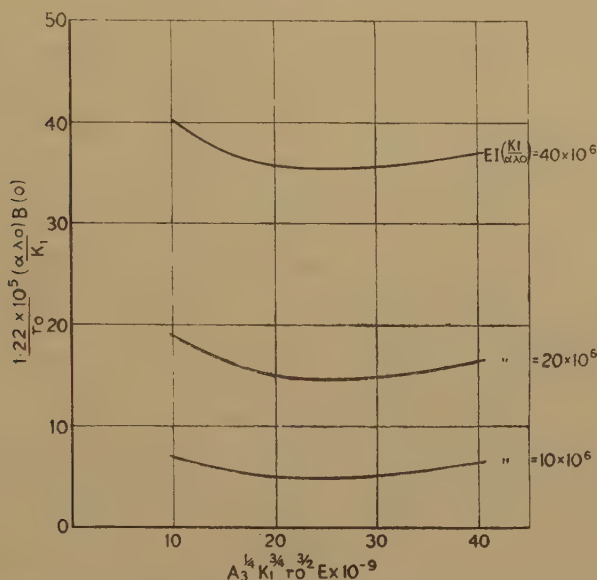
As a particular example of the distribution of radiation brightness across the discharge column given by (10.5) we take the case for which the measured brightness across the arc is shown as the dotted line in fig. 16. For this tube,  $T_0$  was 0.4,  $E$  and  $EI$  approximately  $30 \times 10^8$  and  $30 \times 10^7$  respectively. For these values it is easy to show that  $r_1/r_0 = 0.675$ ,  $T_0 = 6070$ ,  $\beta/2T_w^2 = 19.5$ . Putting these in (10.5) and using fig. 14, we obtain the full line shown in fig. 16. The agreement with the measured



values is so good that it is worth emphasizing that what is calculated here is the total radiation per unit area and what is measured is the candles per unit area. This and the fact that the total radiation is calculated on the assumption of a single mean level, whereas the candles depend in a complicated way upon a number of levels means that, in principle, such close agreement must be fortuitous. There may be reasons nevertheless why the total radiation estimated in this way should give a measure of the candles, but if so, it requires further investigation to establish this agreement.

There is one further property of the total radiation that is easily obtainable and is worth examination. That is the maximum radiation brightness or the radiation per unit area of surface at the peak of the distribution across the arc. In terms of the actual brightness, this

Fig. 17.



would correspond to the brightness measured at the centre of the arc. It is obtainable immediately from (10.3) by putting  $x=0$ .

$$B(0) = \frac{A_3 K_1 \sigma Pr_1}{8\pi} \sqrt{\frac{2}{\beta}} I(T_0). \quad (10.6)$$

To find  $B(0)$ , therefore, we need columns 5, 6 and 11 of the tables;  $I(T_0)$  follows from fig. 14, and we need to interpolate in the tables to find the appropriate values of  $\beta$ . The results are shown in fig. 17. The expected rapid increase with  $EI$  at constant  $E$  is clearly evident; the variation with  $E$  at constant  $EI$  is much smaller and there seem to be regions where an increase of  $E$  actually causes a decrease of  $B(0)$ . As in the case of a similar trend with  $r_1/r_0$ , it is not certain that a similar trend of the actual luminous brightness does not exist, but whether it does or not there is insufficient reason at present to press the correspondence between  $B(0)$  and the actual brightness.

## 11. Conclusion.

It is well perhaps to summarize the results obtained in the light of the basic assumptions made at the outset. In setting up the energy balance equation, two assumptions were made :—

- (a) That the losses other than radiation losses were given by a conduction term proportional to  $T$ .
- (b) That the radiation losses could be accounted for by radiation from a single mean level. This led to a radiation term depending on  $\phi(T)$  given by (1.4), and ignores the effects of absorption and continuous radiation.

In solving the equation the assumptions were made :—

- (c) That the wall temperature  $T$  was constant and the figure of  $1000^\circ \text{K.}$  was assumed thereafter.
- (d) That for the purpose of finding the temperature distribution all the energy was dissipated uniformly within a cylinder of radius  $r_1$ .

It was pointed out in paper A that, whether or not these conditions represented an approximation to those of the high-pressure mercury vapour discharge, it was of considerable interest to study the properties of a system obeying conditions (a) and (b). The fact that in the solution of the energy balance equation three constants appeared, the values of which were found by an appeal to the actual properties of the discharge rather than by direct measurement of the properties of the mercury vapour, made it likely that at least over a limited range of conditions some correspondence between the calculated characteristics and measured properties of the mercury vapour discharge could be expected. The agreement obtained, in fact, appears to be better than would be expected, and it is likely that it would be better still if deductions from the measured quantities could be made with more certainty. For this reason the question arises whether the assumptions made are as drastic as at first sight appears.

So far as (c) is concerned, this anyway would not be likely to cause serious discrepancy with measured values. For the lower values of  $E$  and  $P$ , measurements are usually made in hard glass tubes where the wall temperature cannot be much in excess of  $600\text{--}700^\circ \text{C.}$  On the other hand, if the vapour pressure conditions are to be maintained the wall temperature cannot be very much less than this. In the case of quartz envelopes, which are used for the larger loadings, the wall temperature may be a few hundred degrees higher. Examination of (1.8) and (1.9) shows that even in this case  $T_0$  will be altered by only 1 or 2 per cent. It is true that this effect is not entirely negligible and would modify to some extent the comparison with experiment at the higher pressures. It does not make a major difference, however.

The assumption (d) is more serious. It will affect the temperature distribution and the value of  $T_0$  to a varying extent depending on the

sharpness of the fall-off of temperature from the axis of the discharge. The effect on  $T_0$  is likely to be the greatest error produced in this way since, although  $T_0$  will be less than the actual value, the temperature calculated for other parts of the column will be larger than it should be and, so far as measurable properties of the discharge are concerned, would be likely partially to compensate for the parts where the temperature is too low. The magnitude of the error so introduced could be estimated in two ways, either by taking the distribution found here as a first approximation, substituting in the energy balance equation and solving again to obtain a second approximation; or by taking  $T$  as a function of  $A_3 P^{3/2}/E^2$  different from that assumed in A and finding if an appreciable change in the calculated value of  $T_0$  resulted. It was pointed out in section 4 of A that this choice is largely a matter of intuition and convenience. General physical considerations, however, would suggest that an error introduced by this assumption, while perhaps appreciable, would not by itself be sufficient to destroy correspondence between theory and experiment.

The assumptions (a) and (b) are those which would be expected to prevent agreement with experiments except over a limited range. The conductivity will not, as we have assumed, be proportional to  $T$ ; the effect of the method of choosing the constants we have adopted will be to fit  $\sigma_0 T$  as closely as possible to the actual  $\sigma(T)$ . It is still more certain that the actual radiation function will not be, in detail, anything like the  $\phi(T)$  function we deduce from the assumption of a single level. The actual radiation function will consist mainly of a number of superimposed exponential terms the indices of which will depend on the energy of the level and the coefficients on the transition probabilities to all other levels; this form will be modified by the effects of absorption which means that the population of the various levels will not depend entirely on the temperature. A more serious effect than this, at least at the higher pressures, will be that due to the perturbation of the levels by neighbouring atoms and the consequent change or broadening of the energy levels. This will mean that to some extent at least the indices in the radiation terms are likely to depend on the atom density. Strictly there should also be a term to take account of radiation due to recombination into the various levels. Again this effect will be most marked at the higher states of ionization and at the higher atomic energies. There are other effects also of less importance which should, in a rigorous treatment, be taken into account.

Now, the consequence of all these phenomena will be to give a radiation term not only different, as a function of  $T$ , from the  $\phi(T)$  form assumed here, but which will depend also on other parameters in a complicated manner. The form of the function will almost certainly be one increasing with  $T$  over the whole of the range of  $T$  in which we are interested, but it is possible that its first derivative with respect to  $T$  is not monotonic—the gradual superimposing of the higher levels as  $T$  is increased will be to produce a curve in which the slope changes in an



irregular manner. This function will itself change slowly with the operating conditions of the discharge.

If we were concerned with only a very limited range of operating conditions the method of our choice of constants would be to fit the  $\phi(T)$  curve as closely as possible to the actual curve. Both so far as the radiation is concerned and the thermal conduction term this fit would probably be fairly satisfactory, for any particular state of operation depends on a wide range of temperatures and therefore it is not unreasonable to expect a smoothed form of the functions to serve quite well. As soon, however, as conditions depart from the field in which the constants are chosen, two considerations arise. The first is that the range of temperature in the discharge changes. If the constants are chosen when  $T_0=6000$ , and if we then pass to a condition of operation in which  $T_0=7000$ , in effect we have to extrapolate the mean curve chosen to fit the first condition. The reason why this does not appear to be as serious as it might is because  $T_0$  itself has a limited range (the extreme values for all conditions would seem to be covered by the range  $5500^\circ$  to  $8000^\circ$  K.). Nevertheless, it is one of the most likely reasons for any actual inadequacy of the theory. The second consideration is the one already mentioned in that the actual form of the function will change with the operating conditions. It is impossible to avoid the conclusion that this will have a serious effect at the larger values of  $E$  and  $P$ , and may well account for the discrepancies observed in section 9 in connection with figs. 11 and 12.

The most obvious way to take these effects into account is to introduce another constant by making  $A$  a linear function of the atomic density. This would probably take account partly of absorption as well as of the change of form of the radiation function due to perturbation effects, since absorption itself will, other things being equal, increase with increasing density. This, however, is a purely empirical approach and is somewhat unsatisfactory since the agreement obtained with the present crude theory is such that the introduction of another constant to be chosen from the properties of the discharge would make excellent agreement between theory and experiment very likely. A better method would be to retain the present three constants and attempt to modify the radiation and conduction terms by theoretical considerations, a process which may not prove impossible.

#### *References.*

- (1) Francis, V. J., *Phil. Mag.* xxxvii. p. 433, July (1946).
- (2) Elenbaas, W., *Physica*, ii. p. 757 (1935).
- (3) See Elenbaas, W., *Physica*, ii. p. 169 (1935). Heller, G., *Physics*, vi. p. 389 (1935).
- (4) Elenbaas, W., *Physica*, i. p. 211 (1934).
- (5) Elenbaas, W., *Physica*, ii. p. 787 (1935).



LXIX. *The Limiting Resolving Power of the Electron Microscope.*

By G. LIEBMANN \*†.

[Received March 11, 1946.]

## SUMMARY.

The final resolving power of the magnetic electron microscope is limited by magnetic saturation of the pole pieces which leads to an increase in the minimum focal length and the minimum spherical and chromatic aberration constants if the accelerating voltage is increased. From the minimum spherical aberration constants obtained by combining Ramberg's theoretical work with measurements carried out by the writer, an estimate is obtained of the expected limit of resolving power that may be achieved in electron microscopes of the present type. This limiting resolving power has a slightly marked optimum for an accelerating voltage of 50 KV., and is of the order of 10 to 12 Å.

*Introduction.*

THE resolving power of the electron microscope reached that of the light microscope in 1936, and surpassed it rapidly in the next years. By 1941, several authors reported resolving powers of better than 30 Å., (v. Borries and Ruska <sup>(1)</sup>, Hillier and Vance <sup>(2)</sup>, v. Ardenne <sup>(3)</sup>), the highest resolution claimed at that time being of the order of 22 to 25 Å., although such a high resolution was only attainable with certain types of specimen and under very carefully controlled experimental conditions. The ultimate resolving power achieved experimentally was apparently not improved appreciably beyond this value for some years (Zworykin and Hillier <sup>(4)</sup>, Hillier and Baker <sup>(5)</sup>). Only very recently a higher resolving power was reported by Hillier <sup>(6)</sup>, who used a special, short focus objective.

These results do not bear out the optimistic expectations of some authors. For instance Ruska <sup>(7)</sup> considers that with a focal length of  $f=1.8$  mm. a resolution of 4.5 Å. might be obtained, basing his predictions on the theory developed by Glaser <sup>(8)</sup> and his school.

In this paper, certain consequences of magnetic saturation of the iron pole pieces will be discussed, and it will be seen that the apparent halting of the best experimentally achieved resolution not far below 20 Å is not due to chance, but rather to the inherent limitations of the present type of electron microscope, to which recent designs have already approached rather closely.

\* Communicated by Sir Lawrence Bragg, F.R.S.

† Vacuum Research Laboratory, Cathodeon Ltd., Cambridge.

*Definition of Resolving Power.*

It is known that the resolving power of the electron microscope is limited, even in the absence of any asymmetries, by the diffraction error, and spherical and chromatic aberration. Following Rebsch<sup>(9)</sup> it is usually assumed that the best resolution can be obtained if the radius of the disc of confusion due to spherical aberration

$$\delta_s = C_s \alpha^3 \quad . . . . . (1)$$

equals the radius of the first-order diffraction disc

$$\delta_D = \frac{\lambda}{\alpha}, \quad . . . . . (2)$$

if the object is illuminated by a nearly parallel bundle of electrons of wave-length  $\lambda$ ,  $\alpha$  being the angle of the ray leaving the object towards the objective lens\*.

This leads to the well-known equation

$$\delta_{s \min} = \delta_{D \min} = \lambda (C_s / \lambda)^{\frac{1}{3}}. \quad . . . . . (3)$$

The constant  $C_s$  is the spherical aberration constant of the microscope objective.

The limitation due to chromatic aberration, which is given by

$$\delta_c = (\Delta f)_c \alpha,$$

$(\Delta f)_c$  being the variation of the paraxial focal length  $f$  owing to accelerating voltage or focusing coil current fluctuations, has been frequently disregarded (*e. g.* Ruska<sup>(7)</sup>), or has been considered unimportant compared with the limitation imposed by spherical aberration and diffraction error (Cosslett<sup>(10)</sup>); the numerical example discussed below indicates that this view is not generally justified. The focal length  $f$  depends, even in strong magnetic lenses, only on the ratio  $\left(\frac{V}{H^2}\right)$ ,  $V$  being the accelerating voltage and  $H$  the (maximum or average) magnetic field strength, although the constant of proportionality changes in strong lenses with the parameter  $\left(\frac{V}{H^2}\right)$ . We have, therefore,

$$(\Delta f)_c = C_c \left( \frac{\Delta V}{V} - 2 \frac{\Delta H}{H} \right),$$

$C_c$  being the spherical aberration constant of the lens.

\* It is frequently argued that the scattering of the electrons in the specimen would make this equivalent to a self-luminous body, and that equ. (2) should read  $\delta_D = 0.61 \frac{\lambda}{\alpha}$ . For those specimens which have given so far the highest resolution, viz. heavy metal particles, few of the scattered electrons contribute to the image as they are completely stopped by the specimen, and equ. (2) in its above form appears more appropriate.

If we assume, following Ardenne<sup>(11)</sup>, that the various lens errors combine like error functions, we should write

$$\delta_C = C_C \alpha \sqrt{\left(\frac{\Delta V}{V}\right)^2 + \left(\frac{2\Delta H}{H}\right)^2} \quad (4)$$

as  $\frac{\Delta V}{V}$  and  $\frac{\Delta H}{H}$  will be independent of each other. The superposition of the three main lens errors leads to this minimum radius of the diffraction disc

$$\delta_T = \sqrt{\delta_D^2 + \delta_S^2 + \delta_C^2} \quad (5)$$

The practical limit of resolving power is determined from the least separable distance of two-point objects. The images of two-point objects are still separable if their "discs of confusion" overlap to some extent, although the image contrast will then be reduced; the amount of contrast reduction for a given amount of overlap depends on the radial intensity distribution in the disc of confusion. The term "ultimate resolving power" can, therefore, only have an *approximate meaning*, and the practical limit of resolving power will be better than that determined from the *diameter* of the greatest "disc of confusion," which would give the least distance of two-point objects for which no loss of contrast takes place.

The values  $\delta_D$ ,  $\delta_S$ ,  $\delta_C$  and  $\delta_T$  are, however, the *radii* and not the *diameters* of the respective confusion discs, and it appears legitimate to take  $\delta_T$  as a good approximation to the least distance of two-point objects situated on the optical axis which are separable under practical conditions, making in this way allowance for some overlap of the "images" of the two-point objects, and for the fact that the focus is, in practice, adjusted not for the best focus of the paraxial ray but for the focus giving the best resolution.

$\delta_T$  can be calculated if  $\lambda$ ,  $\frac{\Delta V}{V}$ ,  $\frac{\Delta H}{H}$  and the lens constants  $C_S$  and  $C_C$  are known; if, in particular, the *limits* of the constant  $C_S$  can be ascertained, the limit of resolving power of electron microscopes can be predicted.

#### *Numerical Values of Aberration Constants and of Resolving Power.*

The theoretical determination of the aberration constants has been the subject of intensive study by several investigators, and the results most frequently quoted for magnetic lenses are Glaser's<sup>(8)</sup>. His results are given in terms of a field parameter  $d$  which represents the distance over which the magnetic field has dropped to one-half its maximum value. The application of his results requires, therefore, a knowledge of the actual magnetic field distribution in strong magnetic lenses of small diameter. This knowledge is still fairly inaccurate, and the magnetic field distribution under normal operating conditions is also, as far as is known, somewhat different from the one assumed by Glaser for greater ease of mathematical manipulation.

Another calculation of the aberration constants, which can be more directly applied, is due to Ramberg<sup>(12)</sup>. His calculations refer to the idealised case of a set of cylindrical pole pieces of infinitely great permeability at very small axial distance from each other, but the field distribution assumed by him appears not too far removed from that found in actual lenses. His results are directly related to the internal diameter  $D$  of this "ideal" lens, which corresponds to the bore  $D$  of the pole pieces of actual lenses, and can therefore be compared more easily with the results of measurements. Only three measurements of the spherical aberration constant  $C_s$  of electron microscope lenses have been reported in the previous literature (Ruska<sup>(13)</sup>, Ardenne<sup>(14)</sup>, Dosse<sup>(15)</sup>), but all of them agree reasonably well with Ramberg's calculations. (Dosse's first published value of  $C_s = 0.11$  cm., which appeared to agree with Glaser's more optimistic predictions, was found later to be erroneous, the corrected experimental value being  $C_s = 0.76$  cm.) Measurements of the aberration constants which the writer has carried out on various strong electrostatic and magnetic lenses, and which will be published later, also support Ramberg's theoretical results. It appears, therefore, justified to base a numerical determination of  $\delta_T$  on Ramberg's calculated aberration values.

Before discussing in greater detail the question of the ultimate limit of resolving power, it might be instructive to look at a numerical example. If we assume the following data to be representative of the objective lenses of most modern electron microscopes:

$$V = 50 \text{ KV. } (\lambda = 0.053 \text{ \AA.}),$$

$$D = 0.40 \text{ cm.},$$

$$f = 0.36 \text{ cm.},$$

we should obtain, with  $f/D = 0.85$ , from Ramberg:

$$C_s = 0.90 \text{ cm.}, \text{ and}$$

$$C_c = 0.28 \text{ cm.}$$

Assuming further,  $\frac{\Delta V}{V} = 2 \times 10^{-5}$  and  $\frac{\Delta H}{H} = 3.6 \times 10^{-5}$ , a rather high degree of stabilization, we should obtain from equations (1), (2) and (3):

$$\alpha = 4.9 \times 10^{-3}, \text{ and}$$

$$\delta_D = \delta_s = \delta_c = 10.8 \text{ \AA.}$$

$$\text{This would lead to } \delta_T = 18.6 \text{ \AA.}$$

The resolving power predicted from the aberration data available is therefore very close to that achieved by the best existing instruments of "standard" construction under the most favourable conditions\*.

---

\* Hillier<sup>(6)</sup> reports that only 10–15 exposures out of approximately 25,000 taken with a standard instrument showed a resolution of better than 20 Å.; we are, in our estimates, only concerned with lenses which are of perfect symmetry, and therefore free from astigmatism and coma, and refer therefore to these 10–15 exceptionally good pictures.



To some extent, the lens data assumed may appear arbitrary, but equation (3) indicates that very considerable improvements in the spherical aberration would be needed to improve the resolving power to a marked degree.

It is interesting to note that a reduction of spherical aberration, at the performance reached now, must be accompanied by a reduction of chromatic aberration, as otherwise chromatic aberration would become predominant, even with favourable specimens (*e.g.* metallic colloids).

A fundamental limitation to the stabilization ratio  $\frac{\Delta V}{V}$  is set by the statistical distribution of electron speeds leaving the thermionic cathode, and  $\frac{\Delta H}{H}$  still contains thermal "drift," even if the focusing coils are fed from batteries.

In electron microscopes using electrostatic lenses the part of  $\delta_c$  due to the term  $\frac{\Delta H}{H}$  does not appear, but the constants  $C_s$  and  $C_c$  for electrostatic lenses are greater than those for magnetic lenses of equal bore and focal length, and the ultimate resolving power calculated for an electrostatic electron microscope is therefore somewhat worse than that of a magnetic microscope; the ultimate resolving power is here limited by the maximum electric field strength that can be employed as the minimum electrode spacing, and consequently the focal length, have to be increased with increased accelerating voltage.

### *The Effect of Iron Saturation.*

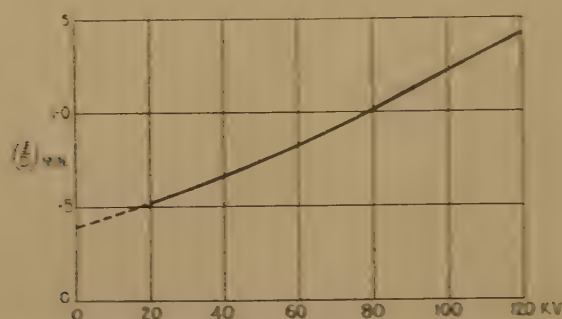
Ramberg's analysis indicates that the best way of reducing the spherical aberration constant  $C_s$  of a magnetic objective lens is a decrease of the focal length  $f$  by an increase in magnetic field strength, as in this case  $C_s \propto f^{-2.7}$ , whereas a mere reduction of the physical lens dimensions would only make  $C_s \propto f^{-1}$ . A limit to the magnetic field strength that can be used is given by the magnetic saturation of the iron pole pieces; as Ramberg's fundamental assumption was an infinitely great permeability of the iron circuit, his investigation cannot give us any information on this important point.

The only measurements from which the minimum value of  $f/D$  as function of accelerating voltage can be determined without further assumption appear to be published by Ardenne<sup>(16)</sup>. Shorter focal lengths were reached by Ruska<sup>(7)</sup>, but the minimum focal length is given by him in absolute values (millimetres), and the physical dimensions of the lens are not given; thus the ratio  $f/D$  cannot be determined.

Fig. 1 shows the value of  $(f/D)_{\min}$  plotted over the accelerating voltage  $V_0$ , which was obtained experimentally by the writer on a typical electron microscope objective of 4.76 mm. bore and 4.76 mm. air gap. The actual measurements were carried out in a special lens test unit over a voltage range  $V_1 = 18$  KV, to 32 KV, and a wide range of magnetic

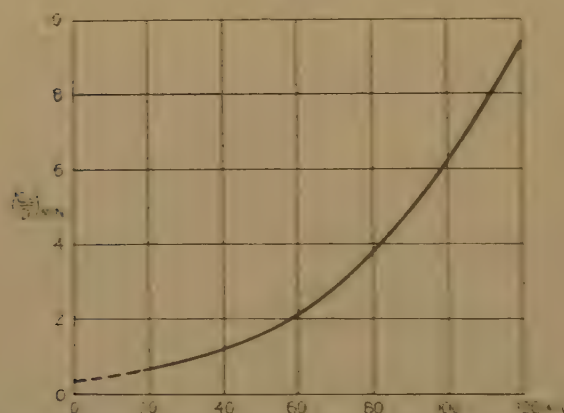
fields  $H_1$ ; the values for the higher accelerating voltages  $V_0$  were obtained from a measured curve of  $(fD)_{\min}$  as function of  $\left(\frac{H_1^2}{V_1}\right)$  by using the relation  $V_0 = \left(\frac{H_{\max}}{H_1}\right)^2 V_1$ . The minimum focal lengths given in fig. 1 are shorter than those of Ardenne, who finds approx.  $(fD)_{\min} = 1.8$  for 100 KV., whereas fig. 1 gives only 1.22. This last value appears nearer

Fig. 1.



Minimum relative focal length  $(fD)_{\min}$  of typical electron microscope objective lens as function of accelerating voltage.

Fig. 2.



Minimum relative spherical aberration constant  $(C_s D)_{\min}$  of typical electron microscope objective as function of accelerating voltage.

to what may be inferred from Ruska's<sup>17</sup> and Dosse's<sup>18</sup> work, being perhaps slightly greater than their values.

We can now use the values of  $(fD)_{\min}$  of fig. 1 to determine  $(C_s D)_{\min}$  as function of accelerating voltage (fig. 2), from Ramberg's calculated  $(\zeta/M)$  values, as

$$(C_s D)_{\min} = (\zeta/M)(fD)_{\min}^2.$$

The spherical aberration parameter  $(C_s/D)_{\min}$  does not depend on the actual physical dimensions of the lens, and is therefore of more general applicability.

From the relation

$$\lambda = 0.384(\text{KV.})^{-\frac{1}{2}}(1 - 5 \times 10^{-4} \text{KV.})$$

and fig. 2 we can determine a parameter

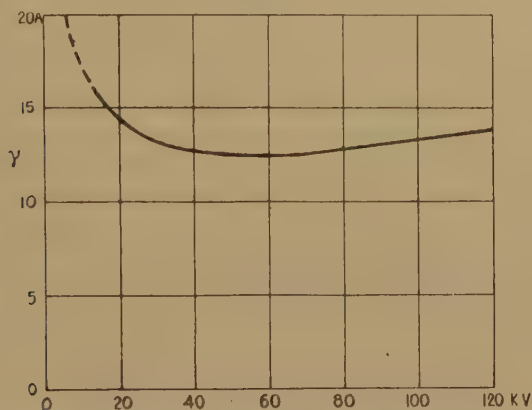
$$\gamma = \lambda^{\frac{2}{3}}(C_s/D)^{\frac{1}{3}} \quad (6)$$

as function of accelerating voltage.

For  $D=1$  cm.,  $\gamma$  would be identical with  $\delta_{s \min}$  (the spherical aberration error for best resolution), as given by equ. (3). For other values of the lens diameter  $D$ , we should find  $\delta_{s \min}$  as

$$\delta_{s \min} = \gamma D^{\frac{1}{3}} \quad (7)$$

Fig. 3.



Resolving power parameter  $\gamma$  (in Angstrom Units) as function of accelerating voltage for typical electron microscope objective. ( $\gamma = \delta_{s \min}$  for  $D=1$  cm.)

The function  $\gamma$  is plotted in fig. 3, the ordinates indicating  $\delta_{s \min}$  in Angstrom units if  $D$  is measured in cm.  $\gamma$  is now a direct measure of the resolving power.

#### *The Limit of Resolving Power.*

The graph (fig. 3) shows the interesting fact that over the mostly used voltage range of 30 to 80 KV. the resolving power under ideal conditions (absence of chromatic aberrations due to differential loss of electron speed in the specimen, and absence of additional lens errors and extraneous disturbances), is *nearly independent of the accelerating voltage*, showing a slightly marked optimum for about 50 to 60 KV. For higher voltages, the resolving power decreases slightly, whereas the decrease is more marked at low accelerating voltages (below 20 KV.). The shape of this curve explains why electron microscopes using very high accelerating voltages (200–300 KV.) give reduced resolving power although the penetration of thick specimens is greatly improved. It also disposes of the

suggestion that operation at very low voltages (10 KV.) may lead to improved resolution.

The plot of  $\gamma$  gives us a fairly close estimate of the best resolving power that may be obtained in future with the present type of magnetic electron microscope. The prospects of increasing appreciably the magnetic saturation limit of the ferro-magnetic materials used now are only slight. Considerable variations in the constructional features of the pole pieces showed that it is possible to influence the properties of magnetic lenses to a minor degree by modifying the shape and distances of the pole pieces (*e.g.*, the focal length decreases with reduced gap length), but that for constant voltage the aberration parameter  $(C_s/D)_{\min}$  is only very slightly affected by such modifications. It appears unlikely to the writer that such improvements in  $(C_s/D)_{\min}$  can be obtained by better design of pole pieces that a *substantially* better resolving power than that given in fig. 3 could be produced, because the resultant resolving power would only be proportional to the fourth root of the improvement.

In the light of the past experience of several investigators, it would seem reasonable to assume that the smallest electron microscope objective lens which could be made with a sufficiently high degree of accuracy would have a bore of  $D=0.10$  to  $0.20$  cm. The combination of the three principal aberrations  $\delta_D$ ,  $\delta_s$  and  $\delta_C$  would make  $\delta_T > \sqrt{2} \delta_s$ , and very likely, with such voltage and current stabilization values as might be feasible, rather nearer  $\sqrt{3} \delta_s$ . Taking these two facts together, we can expect that our curve, fig. 3, will be very nearly a direct graph in Angstroms of the ultimate resolving power of which the present type of electron microscope might be capable under the most favourable conditions.

Very recently, Hillier<sup>(6)</sup> succeeded in obtaining a resolving power of 13 Å. to 15 Å., if the resolving power is defined as it was in the present paper\*, using a specially designed objective lens of 2.5 mm. focal length which was particularly free from axial asymmetries and was operated at the optimum voltage of 50 KV. This appears to be in good agreement with our predictions.

Under ordinary conditions, the ultimate resolving power will be rather worse owing to additional lens errors, as coma and astigmatism, and owing to additional chromatic aberration due to the specimen itself.

The great improvement in resolving power envisaged by some previous authors appears not realizable with the present type of electron microscope owing to its inherent limitations, and the writer agrees with Gabor's<sup>(17)</sup> recent opinion that a considerable advance in resolving power would require the discovery of entirely new means of correcting electron lenses.

---

\* Hillier's definition of resolving power is based on the minimum distance of two-point objects that can be *just* resolved. According to this definition the resolving power is approximately  $\frac{2}{3}\delta_T$ ,  $\delta_T$  being given by our equ. (5), and the predicted limiting resolving power would be about 9 Å., whereas Hillier's experimental mean value is 10.5 Å.



### Acknowledgments.

The writer wishes to thank Sir Lawrence Bragg for the interest he has taken in this paper, and Dr. V. E. Cosslett and Dr. D. Gabor for stimulating discussions on the question of the ultimate resolving power of electron microscopes.

*Note added in proof.*—Whilst this manuscript was in the press, a paper by Cosslett <sup>(18)</sup> was published, in which the same problem is discussed on similar lines and with almost identical results. Cosslett approaches the problem more from the theoretical side, and his investigation is therefore complementary to the writer's, which is mainly based on *measurements* of lens performance.

### References.

- (1) B. v. Borries and E. Ruska, *Erg. exact. Naturw.* xix. p. 271 (1940).
- (2) J. Hillier and A. W. Vance, *Proc. I. R. E.* xxix. p. 167 (1941).
- (3) M. v. Ardenne, *Phys. Z.* xlii. p. 72 (1941).
- (4) V. K. Zworykin and J. Hillier, *Phys. Rev.* lxiv. p. 315 (1943).
- (5) J. Hillier and R. F. Baker, *Journ. Appl. Phys.* xvi. p. 469, fig. 2 (1945).
- (6) J. Hillier, *Journ. Appl. Phys.* xvii. p. 307 (1946).
- (7) E. Ruska, *Arch. f. Elektrot.* xxxvi. p. 431 (1942).
- (8) W. Glaser, *Z. f. Phys.* cxvii. p. 285 (1941).
- (9) R. Rebsch, *Ann. d. Phys.* xxxi. p. 551 (1938).
- (10) V. E. Cosslett, *Journ. Sci. Inst.* xxii. p. 170 (1945).
- (11) M. v. Ardenne, *Z. f. Phys.* cviii. p. 338 (1938).
- (12) E. G. Ramberg, *Journ. Appl. Phys.* xiii. p. 582 (1942).
- (13) E. Ruska, *Z. f. Phys.* lxxxix. p. 90 (1934).
- (14) M. v. Ardenne, *Z. f. tech. Phys.* xx. p. 289 (1939).
- (15) J. Dosse, *Z. f. Phys.* cxvii. p. 722 (1941) and *Z. f. Phys.* cxviii. p. 375 (1941).
- (16) M. v. Ardenne, *Z. f. Phys.* cxvii. p. 602 (1941).
- (17) D. Gabor, 'The Electron Microscope,' Chap. 12. (London, Hulton Press Ltd., 1945).
- (18) V. E. Cosslett, *Proc. Phys. Soc.* lviii p. 443 (1946).

LXX. *The Effect of Fluctuation Noise Interference on Pulse Distortion.*

By P. J. HILTON, B.A.\*

[Received November 14, 1945.]

### SUMMARY.

This paper deals with the effect of the superposition of fluctuation noise on a distorted signal pulse. The probability distribution of the additional distortion of the pulse is discussed and tables are given for its ready evaluation.

### Introduction.

FLUCTUATION noise may be due either to the thermal motion of electrons in conductors or to the granular structure of the plate current of a

\* Communicated by H. J. Josephs,

vacuum tube. These types of noise are known respectively as thermal-agitation noise and shot-effect noise.

Since the distribution of the instantaneous value of the noise envelope is known<sup>(1)</sup>, it is possible to extend the analysis to the consideration of noise interference with a signal pulse in a transmission system. The distortion so produced is measured by the difference in the duration of the pulse with and without the superposition of fluctuation noise. The duration of the pulse is taken as the difference between the times at which the envelope of the pulse passes through half its maximum amplitude during the building-up and the dying-away periods of the pulse.

For the case when the signal and noise are confined by means of selective circuits to a narrow frequency band, the relationship between noise amplitude and the consequent additional pulse distortion has been investigated elsewhere<sup>(2)</sup>. Confining attention to one end of the pulse, it may be shown that when the ratio  $a$  of the noise amplitude  $A$  to the peak value of the amplitude of the signal voltage  $V$  is considerably less than unity, the distortion  $D$  at one end of the pulse is given by the simple expression

$$D = (a \cos \theta)t, \quad . . . . . (i)$$

where  $\theta$  is the angle between the noise and signal voltage, and  $t$  is the time that would be taken for the pulse to reach its maximum value at the rate applying at the half-amplitude point. In (i),  $D$  is the difference in the time taken for the pulse to reach half its maximum value with and without the superposition of fluctuation noise, so that, if  $D'$  is the distortion at the other end of the pulse, the total distortion of the pulse  $D_T$  is given by

$$D_T = D + D' \quad . . . . . (ii)$$

It is the object of this paper to obtain the probability distributions of  $D$  and  $D_T$ . The distribution of  $D$  is obtained from (i) by using a theorem, due to E. V. Huntington<sup>(3)</sup>, for the distribution of the product of two variates. This theorem, though often implicit in statistical work, is seldom explicitly stated. An expression is also derived for the probability integral associated with  $D$ ; the difficulty encountered in evaluating this integral by classical methods led to the discovery of a very simple transformation which converted the integral into a form suitable for treatment by the Heaviside operational method. The transformation consists of the introduction of an arbitrary independent variable into the integrand, and subsequently dropping this variable when it has fulfilled its purpose. Such tactics were never, so far as is known, used by Heaviside himself. This manœuvre was later substantiated by obtaining the same result with the aid of Parseval's theorem of Fourier Transforms.

The result, giving the distribution of the additional distortion at one end of the pulse, is readily extended to cover the problem of the total

distortion at both ends, provided the noise amplitudes at the two ends can be considered independent. In this case, a workable result, subject to the approximation involved in (i), is obtained and tabulated.

### The Distribution of D.

From (i) we have

$$D = (a \cos \theta)t \\ = \frac{(A \cos \theta)t}{V}.$$

$$\therefore \left| \frac{DV}{t} \right| = A |\cos \theta| = A |y|, \text{ say.}$$

We now quote the above-mentioned theorem due to Huntington, which states that if  $f(x)$  is the distribution of  $x$ , such that  $\int_0^\infty f(x)dx=1$ , and  $F(y)$  is the distribution of  $y$ , such that  $\int_0^\infty F(y)dy=1$ , and  $\phi(u)$  is the distribution of  $u$ , such that  $\int_0^\infty \phi(u)du=1$ , and if  $u=xy$ , then

$$\phi(u) = \int_0^\infty f(x) F\left(\frac{u}{x}\right) \frac{dx}{x}. \quad \dots \dots \dots \text{(iii)}$$

We require to express the probability distributions of  $A$  and  $|y|$  as  $f(A)$  and  $F(|y|)$ .

Now  $f(A)$  is known<sup>(1)</sup>; its value is

$$\frac{A}{E^2} e^{-A^2/2E^2},$$

where  $E$  is defined as the R.M.S. value of the noise voltage.

Again, since  $\theta$  may take all values between  $0$  and  $\pi$  with equal likelihood, the probability distribution of  $y$  is

$$\frac{1}{\pi \sqrt{1-y^2}}.$$

Therefore the probability distribution of  $|y|$  is

$$\frac{2}{\pi \sqrt{1-y^2}},$$

such that

$$\int_0^1 \frac{2dy}{\pi \sqrt{1-y^2}} = 1.$$

Thus

$$F(|y|) = \frac{2}{\pi \sqrt{1-y^2}}, \quad 0 \leq y \leq 1 \\ = 0, \quad y > 1.$$

Consequently, the distribution of

$$u \left( \equiv \left| \frac{DV}{t} \right| \right) \text{ is } \phi(u),$$

3 E 2

where

$$\phi(u) = \int_0^\infty \frac{A}{E^2} e^{-A^2/2E^2} F\left(\frac{u}{A}\right) \frac{dA}{A}$$

and

$$F\left(\frac{u}{A}\right) = \frac{2A}{\pi\sqrt{A^2-u^2}}, \quad A \geq u \\ = 0, \quad A < u.$$

$$\therefore \phi(u) = \int_u^\infty \frac{A}{E^2} e^{-A^2/2E^2} \frac{2A}{\pi\sqrt{A^2-u^2}} \frac{dA}{A}.$$

Make the substitution  $B^2 = A^2 - U^2$ , so that  $A dB = B dB$ . Then

$$\phi(u) = \int_0^\infty \frac{1}{E^2} \frac{e^{-u^2/2E^2} e^{-B^2/2E^2} \cdot 2B dB}{\pi B} \\ = \sqrt{\frac{2}{\pi}} \frac{e^{-u^2/2E^2}}{E},$$

since

$$\sqrt{\frac{2}{\pi}} \frac{1}{E} \int_0^\infty e^{-B^2/2E^2} dB = 1.$$

From physical considerations it is apparent that  $D$  is distributed symmetrically about zero. It therefore follows from the above that

$$\frac{DV}{t}$$

is distributed normally with zero mean and standard deviation  $E$ , and the probability that the modular value of the distortion exceeds  $|D_0|$  is given by

$$P_{D_0} = \int_{U_0}^\infty \sqrt{\frac{2}{\pi}} \frac{e^{-u^2/2E^2}}{E} du \\ = \operatorname{erfc} \frac{U_0}{\sqrt{2}E}, \quad \dots \dots \dots \text{(iv)}$$

where  $\operatorname{erfc}(x)$  is the complement of the error function, and

$$U_0 \left( \equiv \left| \frac{D_0 V}{t} \right| \right)$$

is the noise amplitude corresponding to a distortion of  $D_0$ , when the angle between the noise and signal voltage is zero; that is to say,  $U_0$  is the noise amplitude producing a maximum distortion of  $D_0$ .

#### Probability Integral.

If  $A$  is regarded as constant, the probability of the modular value of the distortion exceeding  $|D_0|$  is given by

$$P_{D_0} = \frac{2}{\pi} \cos^{-1} \frac{|D_0 V|}{At} = \frac{2}{\pi} \cos^{-1} \frac{U_0}{A}.$$



However, since the probability,  $dp_A$ , that the instantaneous value envelope lies between  $A$  and  $A+dA$  is given by

$$dp_A = \frac{A}{E^2} e^{-A^2/2E^2} dA,$$

it follows that the required expression for  $P_{D_0}$ , defined as in the previous section, is

$$\begin{aligned} P_{D_0} &= \int_{D=D_0}^{D=\infty} \frac{2A}{\pi E^2} e^{-A^2/2E^2} \cos^{-1} \left( \frac{U_0}{A} \right) dA \\ &= \int_{U_0}^{\infty} \frac{2A}{\pi E^2} e^{-A^2/2E^2} \cos^{-1} \frac{U_0}{A} dA. \quad \dots \quad (v) \end{aligned}$$

To evaluate this integral, we first integrate by parts, getting

$$\begin{aligned} P_{D_0} &= - \frac{2}{\pi} e^{-A^2/2E^2} \cos^{-1} \left( \frac{U_0}{A} \right) \Big|_{U_0}^{\infty} + \int_{U_0}^{\infty} \frac{2}{\pi} e^{-A^2/2E^2} \frac{U_0}{A \sqrt{A^2 - U_0^2}} dA \\ &= \frac{2}{\pi} \int_{U_0}^{\infty} e^{-A^2/2E^2} \frac{U_0}{A \sqrt{A^2 - U_0^2}} dA. \end{aligned}$$

Now make the substitution  $A^2 = U_0^2 + 2E^2x^2$ , so that  $A dA = 2E^2x dx$ .

Then

$$P_{D_0} = \frac{2}{\pi} \int_0^{\infty} e^{-U_0^2/2E^2} e^{-x^2} \frac{2E^2 U_0 x}{U_0^2 + 2E^2 x^2} \frac{1}{\sqrt{2E^2 x}} dx,$$

or, putting  $\lambda = U_0^2/2E^2$ ,

$$P_{D_0} = \frac{2\sqrt{\lambda}e^{-\lambda}}{\pi} \int_0^{\infty} \frac{e^{-x^2}}{x^2 + \lambda} dx. \quad \dots \quad (vi)$$

At this stage the transformation is made which converts (vi) into operational form. The classical derivation of the solution of this integral will also be given.

We consider the integral

$$I \equiv \int_0^{\infty} \frac{e^{-x^2 t}}{x^2 + \lambda} dx,$$

so that

$$\lim_{t \rightarrow 1} \frac{2\sqrt{\lambda}e^{-\lambda}}{\pi} I = P_{D_0}. \quad \dots \quad (vii)$$

Now we know that the operational equivalent of  $e^{-at}$  is

$$\frac{p}{p+\alpha}, \text{ where } p = \frac{d}{dt}.$$

Consequently we may write

$$I = \int_0^{\infty} \frac{p dx}{(p+x^2)(\lambda+x^2)}.$$

Taking partial fractions,

$$\begin{aligned} I &= \frac{p}{p-\lambda} \int_0^\infty \left( \frac{1}{\lambda+x^2} - \frac{1}{p+x^2} \right) x \\ &= \frac{p}{p-\lambda} \left( \frac{1}{\sqrt{\lambda}} \tan^{-1} \frac{x}{\sqrt{\lambda}} - \frac{1}{\sqrt{p}} \tan^{-1} \frac{x}{\sqrt{p}} \right) \Bigg|_0^\infty \\ &= \frac{\pi}{2} \frac{p}{p-\lambda} \left( \frac{1}{\sqrt{\lambda}} - \frac{1}{\sqrt{p}} \right) \\ &= \frac{\pi}{2} \frac{1}{\sqrt{\lambda}} \frac{1}{1 + \sqrt{\frac{\lambda}{p}}}. \end{aligned}$$

Expanding in inverse powers of  $\sqrt{p}$ , we have

$$I = \frac{\pi}{2} \frac{1}{\sqrt{\lambda}} \left( 1 - \frac{\lambda^{\frac{1}{2}}}{p^{\frac{1}{2}}} + \frac{\lambda}{p} - \frac{\lambda^{\frac{3}{2}}}{p^{\frac{3}{2}}} + \dots \right).$$

Thus far in the analysis  $p$  has been regarded as a constant, for purposes of integration. At this stage, however, we make the substitution

$$\frac{1}{p^n} = \frac{t^n}{\Gamma(n+1)},$$

where  $\Gamma(n+1)$  is Euler's Gamma function and equals  $n!$  for  $n$  an integer.

We then have 
$$I = \frac{\pi}{2} \frac{1}{\sqrt{\lambda}} \left( e^{\lambda t} - \sum_{n=0}^{\infty} \frac{(\lambda t)^{n+\frac{1}{2}}}{\Gamma(n+\frac{3}{2})} \right).$$

Putting  $t=1$ , and using (vii), we have

$$\begin{aligned} P_{D_0} &= \frac{\pi}{2} \frac{1}{\sqrt{\lambda}} \left( e^{\lambda} - \sum_{n=0}^{\infty} \frac{\lambda^{n+\frac{1}{2}}}{\Gamma(n+\frac{3}{2})} \right) \frac{2\sqrt{\lambda}e^{-\lambda}}{\pi} \\ &= 1 - e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^{n+\frac{1}{2}}}{\Gamma(n+\frac{3}{2})}. \quad \dots \dots \dots \text{(viii)} \end{aligned}$$

Now consider

$$f(\mu) \equiv e^{-\mu^2} \sum_{n=0}^{\infty} \frac{\mu^{2n+1}}{\Gamma(n+\frac{3}{2})},$$

so that  $\lambda = \mu^2$ , and  $f(0) = 0$ .

Differentiating w.r.t  $\mu$ ,

$$\begin{aligned} f'(\mu) &= e^{-\mu^2} \sum_{n=0}^{\infty} \left[ \frac{(2n+1)\mu^{2n}}{\Gamma(n+\frac{3}{2})} - \frac{2\mu^{2n+2}}{\Gamma(n+\frac{3}{2})} \right] \\ &= e^{-\mu^2} \frac{2}{\sqrt{\pi}}. \end{aligned}$$

$$\therefore \int_0^\mu f'(\mu) d\mu = \int_0^\mu \frac{2}{\sqrt{\pi}} e^{-\mu^2} d\mu,$$

$$\therefore f(\mu) = \text{erf } \mu,$$

$$\begin{aligned}\therefore P_{D_0} &= 1 - \operatorname{erf} \sqrt{\lambda} \\ &= \operatorname{erfc} \sqrt{\lambda} \\ &= \operatorname{erfc} \frac{U_0}{\sqrt{2E}}, \text{ which agrees with (iv).}\end{aligned}$$

It should be noted that in the course of this analysis, two interesting results have been established. In the first place, we have obtained a convergent expansion for  $e^{\mu^2} \operatorname{erf} \mu$ , viz.,

$$e^{\mu^2} \operatorname{erf} \mu = \sum_{n=0}^{\infty} \frac{\mu^{2n+1}}{\Gamma(n+\frac{3}{2})} \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{ix})$$

Second, it has been shown that the operational equivalent of

$$e^{\lambda t} \operatorname{erfc} \sqrt{\lambda t} \text{ is } \frac{1}{1 + \sqrt{\frac{\lambda}{\rho}}} \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{x})$$

An alternative procedure from (vi) is to make use of Parseval's theorem of Fourier transforms, which states that if  $F(x)$  is the Fourier transform of  $f(x)$  and  $G(x)$  the Fourier transform of  $g(x)$ , then

$$\int_{-\infty}^{\infty} F(x)G(x)dx = \int_{-\infty}^{\infty} f(x)g(-x)dx \cdot \cdot \cdot \cdot \cdot \quad (\text{xi})$$

In the above,  $F(x)$  is said to be the Fourier transform of  $f(x)$  if

$$F(t) = \int_{-\infty}^{\infty} f(x)e^{2\pi ixt}dx, \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{xii})$$

where  $i = \sqrt{-1}$ .

Now the Fourier transform of  $e^{-x^2}$  is  $\sqrt{\pi}e^{-\pi^2x^2}$  and the Fourier transform of  $\frac{1}{x^2 + \lambda}$  is  $\frac{\pi}{\sqrt{\lambda}}e^{-2\pi\sqrt{\lambda}|x|}$

$$\therefore \int_{-\infty}^{\infty} \frac{e^{-x^2}}{x^2 + \lambda} dx = \int_{-\infty}^{\infty} \frac{(\pi)^{\frac{3}{2}}}{\sqrt{\lambda}} e^{-(\pi^2x^2 + 2\pi\sqrt{\lambda}|x|)} dx,$$

$$\begin{aligned}\therefore P_{D_0} &= \frac{2\sqrt{\lambda}e^{-\lambda}}{\pi} \int_0^{\infty} \frac{e^{-x^2}}{x^2 + \lambda} dx \\ &= \frac{\sqrt{\lambda}e^{-\lambda}}{\pi} \int_{-\infty}^{\infty} \frac{e^{-x^2}}{x^2 + \lambda} dx,\end{aligned}$$

since

$$\frac{e^{-x^2}}{x^2 + \lambda}$$

is an even function of  $x$ . Consequently, applying Parseval's theorem,

$$\begin{aligned}P_{D_0} &= \frac{\sqrt{\lambda}e^{-\lambda}}{\pi} \int_{-\infty}^{\infty} \frac{(\pi)^{\frac{3}{2}}}{\sqrt{\lambda}} e^{-(\pi^2x^2 + 2\pi\sqrt{\lambda}|x|)} dx \\ &= 2e^{-\lambda}\sqrt{\pi} \int_0^{\infty} e^{-(\pi^2x^2 + 2\pi\sqrt{\lambda}x)} dx,\end{aligned}$$

since  $e^{-(\pi^2 x^2 + 2\pi\sqrt{\lambda}|x|)}$  is an even function of  $x$ . Now introduce the substitution  $\pi x + \sqrt{\lambda} = y$ , so that  $\pi^2 x^2 + 2\pi\sqrt{\lambda}x = y^2 - \lambda$ .

Then

$$\begin{aligned} P_{D_0} &= 2e^{-\lambda}\sqrt{\pi} \int_{\sqrt{\lambda}}^{\infty} e^{-y^2} e^{\lambda} \frac{dy}{\pi} \\ &= \frac{2}{\sqrt{\pi}} \int_{\sqrt{\lambda}}^{\infty} e^{-y^2} dy \\ &= \operatorname{erfc} \sqrt{\lambda}, \end{aligned}$$

which agrees with the result obtained by the operational transformation.

### *Distortion at Both Ends of the Pulse.*

To obtain the distribution of the total distortion at both ends of the pulse we employ a familiar result in the mathematical theory of statistics, namely, that the sum of samples of  $N$  drawn from a normal population, of mean  $\bar{X}$  and standard deviation  $\sigma$ , is itself normally distributed with mean  $N\bar{X}$  and standard deviation  $\sigma\sqrt{N}$ .

In our case,  $N=2$ ,  $\bar{X}=0$ ,  $\sigma=E$ . It therefore follows that the total distortion is distributed normally about zero mean with standard deviation  $E\sqrt{2}$ .

Consequently the probability of the total additional distortion exceeding  $D_T$  is given by

$$\begin{aligned} P_{D_T} &= \int_{U_T}^{\infty} \sqrt{\frac{2}{\pi}} \frac{1}{E\sqrt{2}} e^{-x^2/2 \cdot 2E^2} dx \\ &= \int_{U_T}^{\infty} \frac{1}{\sqrt{\pi}E} e^{-x^2/4E^2} dx \\ &= \operatorname{erfc} \frac{U_T}{2E}, \quad \dots \dots \dots (ziii) \end{aligned}$$

where  $U_T$  is the noise amplitude producing a maximum distortion of  $D_T$  at one end of the pulse.

### *Conclusion.*

From the preceding analysis the following table, giving  $P_{D_0}$  and  $P_{D_T}$  as functions of  $\frac{A}{E}$ , the noise peak ratio, has been calculated. An experimental investigation of the general problem of the effect of fluctuation noise interference on pulse distortion is being made at the Post Office Research Station. The author wishes to thank Dr. W. G. Radley, Controller of Research, for permission to publish this paper.



$\frac{A}{E} \left( = \frac{U_0}{E} \text{ or } \frac{U_T}{E} \right)$	$P_{D_0}$	$P_{D_T}$	$\frac{U_T}{E}$	$P_{D_T}$
0.5	.6171	.7237	3.6	.0109
0.6	.5485	.6714	3.7	.0089
0.7	.4839	.6206	3.8	.0072
0.8	.4237	.5716	3.9	.0058
0.9	.3681	.5245	4.0	.0047
1.0	.3173	.4795	4.1	.0037
1.1	.2713	.4367	4.2	.0030
1.2	.2301	.3961	4.3	.0024
1.3	.1936	.3580	4.4	.0019
1.4	.1615	.3222	4.5	.0015
1.5	.1336	.2888	4.6	.0012
1.6	.1096	.2579	4.7	.0009
1.7	.0891	.2293	4.8	.0007
1.8	.0719	.2031	4.9	.0005
1.9	.0574	.1791	5.0	.0004
2.0	.0455	.1573	5.5	.0001
2.1	.0357	.1376	6.0	.00002
2.2	.0278	.1198	6.5	.000004
2.3	.0214	.1039	7.0	.0000007
2.4	.0164	.0897		
2.5	.0124	.0771		
2.6	.0093	.0660		
2.7	.0069	.0562		
2.8	.0051	.0477		
2.9	.0037	.0403		
3.0	.0027	.0339		
3.1	.0019	.0284		
3.2	.0014	.0237		
3.3	.0010	.0196		
3.4	.0007	.0162		
3.5	.0005	.0133		
4.0	.00006	.0047		
5.0	.0000006	.0004		
6.0	.000000002	.00002		

In the above table,  $|D_0| = \frac{U_0 t}{V}$ ,  $|D_T| = \frac{U_T t}{V}$ , and A may be read as either  $U_0$  or  $U_T$ .

#### References.

- (1) Discussion on "The Distribution of Amplitude with time in Fluctuation Noise," by Vernon D. Landon, Proc. I. R. E., p. 426 (Sept. 1942).
- (2) "The Probability Distribution of Distortion Magnitudes due to Inter-channel Interference in Multi-Channel Pulse Transmission Systems," by Dr. D. G. Tucker, J. I. E. E. (to be published).
- (3) "Frequency Distribution of Product and Quotient," by E. V. Huntington, Annals of Mathematical Statistics, No. 10 (1939).

LXXI. "*Lagrangean*" Formulæ for the Direct Calculation of Harmonic Output and of Intermodulation Products.

By A. BLOCH, Dr.-Ing., M.Sc.

(Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England)\*.

[Received May 9, 1946.]

The usual formulæ for the calculation of the harmonic output of a valve, based on a knowledge of the anode current for a number of equal steps of grid potential, always assume that the amplitude of the grid oscillation is an integral multiple of these steps (Ref. <sup>(1)</sup>), the only set of formulæ which so far catered for non-integral values was that of Espley (Ref. <sup>(2)</sup>), based on the use of a difference table †. This latter method was recently extended by the author to allow the calculation of intermodulation products (Ref. <sup>(3)</sup>); it was pointed out then that this method—being based on an interpolating curve which passes exactly through all the given current ordinates, was theoretically superior to another method for the calculation of these intermodulation products (Ref. <sup>(4)</sup>) which used, for the purpose of interpolation, a series of interlaced parabolæ of lower order. The advantage of this latter method, however, was that it required a minimum of arithmetical work: all that was necessary was the multiplication of the given current ordinates by predetermined coefficients and the addition of these products. ("Lagrangean" type of formulæ.)

The purpose of this short note is to show how it is possible in a quite general way to convert results obtained by the difference method into formulæ of the Lagrangean type; this conversion will be shown hereafter, first for the case of a single input frequency (harmonic output) and then for the case of two different input frequencies (intermodulation products) ‡.

### 1. Harmonic Output for a Single Input Frequency.

It will, perhaps, be welcome if the principles of Espley's method are shortly outlined again.

If a difference table is formed from the given values of anode current, in particular, if the line of central differences is calculated, we have thereby defined implicitly (*e. g.*, on the basis of the Newton-Stirling interpolation

---

\* Communicated by Sir Clifford Paterson, F.R.S.

† It is supposed here that the magnitude of the grid voltage steps is unalterably fixed by the manner in which the valve characteristic is given. It is, of course, possible to obtain additional ordinates by a process of interpolation; however, if this is to do full justice to the available data, simple linear interpolation may be ruled out.

‡ Dr. Espley has informed me that he actually used such conversion in the form shown in Ref. <sup>(2)</sup> to obtain the coefficients given in Ref. <sup>(1)</sup>.

formula) a polynomial which assumes exactly the given current values for the stated values of grid potential. If we assume the grid potential to vary as  $a \cdot \cos \omega_a t$ , it would only be necessary to substitute this value into the polynomial and obtain thereby, after some transformations and ordering, all the various components which make up the valve output under these conditions. The magnitude of these components is thus determined by the coefficients of the polynomial; as these in turn are determined by the line of central differences, Espley was able to show how they can be calculated directly from this line of central differences.

If we denote by  $C_p$  the magnitude of the output component of frequency  $p\omega_a$  ( $p=0, 1, 2 \dots$ ), in particular by  $C_p(0)$ , the magnitude of this component when the working point of the valve coincides with  $i_0$ , the central value of the difference table, and by  $F_m^p(a)$ , certain functions of the input amplitude as given in Espley's paper\*, all of his results are of either of the following two forms

$$C_p(0) = \sum_{k=0}^{K} F_{2k}^p(a) \times \delta^{2k} i_0 \text{ for } p \text{ even,} \quad . . . . . (1)$$

or 
$$C_p(0) = \sum_{k=0}^{K-1} F_{2k+1}^p(a) \times \mu \delta^{2k+1} i_0, \quad . . . . . (2)$$

where  $\mu$  and  $\delta$  are Sheppard's operators<sup>(5)</sup> and where  $2K+1$  is the number of entries to the difference table.

Now, all we have to do for our present purpose is to express the differences  $\delta^{2k} i_0$  or  $\mu \delta^{2k+1} i_0$  in terms of the original entries  $i_{-n} \dots i_{+n}$  of the difference table. This transformation can be carried out very simply by the use of the operator  $E$  ( $E i_0 = i_1$ ;  $E^w i_0 = i_w$ ).

By definition we have

$$\delta = E^{\frac{1}{2}} - E^{-\frac{1}{2}}, \quad . . . . . (3)$$

hence

$$\begin{aligned} \delta^{2k} &= (E^{\frac{1}{2}} - E^{-\frac{1}{2}})^{2k} = E^k - C_1^{2k} \cdot E^{k-1} + \dots \\ &= \sum_{t=0}^{t=2k} (-1)^t C_t^{2k} \cdot E^{k-t} \quad . . . . . (4) \end{aligned}$$

Furthermore, by definition

$$\mu = \frac{1}{2}(E^{\frac{1}{2}} + E^{-\frac{1}{2}}), \quad . . . . . (5)$$

hence

$$\begin{aligned} \mu \delta &= \frac{1}{2}(E^{\frac{1}{2}} + E^{-\frac{1}{2}}) \times (E^{\frac{1}{2}} - E^{-\frac{1}{2}}) \\ &= \frac{1}{2}(E^1 - E^{-1}). \quad . . . . . (6) \end{aligned}$$

Therefore

$$\begin{aligned} \mu \delta^{2k+1} &= \mu \delta \delta^{2k} = \frac{1}{2}(E^1 - E^{-1}) \times \sum_{t=0}^{t=2k} (-1)^t C_t^{2k} E^{k-t} \\ &= \frac{1}{2} \sum_{t=0}^{t=2k} (-1)^t C_t^{2k} \times [E^{k-t+1} - E^{k-t-1}]. \quad . . . . . (7) \end{aligned}$$

---

\* The notation here employed is that used in Ref. (3).

Dealing first with the case  $p$  even, we have from (1)

$$\begin{aligned} C_p(0) &= \sum_{k=0}^{k=K} F_{2k}^p(a) \times \sum_{t=0}^{t=2k} (-1)^t C_t^{2k} E^{k-t} \times i_0 \\ &= \sum_{k=0}^{k=K} \sum_{t=0}^{t=2k} F_{2k}^p(a) (-1)^t C_t^{2k} E^{k-t} \times i_0. \quad \dots \quad (8) \end{aligned}$$

For the purpose of evaluation it is best to write this double sum in the form of a rectangular array, say with  $k$  denoting the column and  $t$  denoting the row occupied by each element. The factors  $F_{2k}^p(a)$  being constant for each column can be extracted and placed at the top of each column. Similarly, the factors  $E^{k-t}$  are constant for each diagonal sloping down to the right and need not be written. The array consists, then, simply of the binomial coefficients  $C_t^{2k}$  of appropriate signs, as illustrated in Table I.

TABLE I.

	0	0	2	3	—————> $k$		
	$F_0$	$F_2$	$F_4$	$F_6$	.	.	.
0	1	1	1	1	.	.	.
1		-2	-4	-6	.	.	.
2		1	6	15	.	.	.
3			-4	-20	.	.	.
4			1	15	.	.	.
$\downarrow$ $t$				-6	.	.	.
				1	.	.	.

If we order the double sum according to the powers of  $E^{k-t} = E^w$ , we get then

$$C_p(0) = \sum_{w=-K}^{w=+K} E^w i_0 \times L_p^w(a, K) = \sum_{w=-K}^{w=+K} i_w \times L_p^w(a, K), \quad \dots \quad (9)$$

where  $L_p^w(a, k)$  denotes the sum of the appropriate diagonal of the array. The evaluation of these sums actually completes our task, for equation (9) is a formula of the desired form.

In general terms we have, on account of  $t = k - w$

$$L_p^w(a, K) = \sum_{k=0}^{k=K} (-1)^{k-w} C_{k-w}^{2k} F_{2k}^p(a), \quad \dots \quad (10)$$

where, of course, all those coefficients  $C_{k-w}^{2k} = 0$  where  $k - w < 0$ .





## 2. Harmonic Output for Two Input Frequencies (Intermodulation Products).

If a signal comprising two frequencies  $\omega_a$  and  $\omega_b$  is applied to the grid of the valve the output is of the form

$$i_a = \sum_p \sum_q C_{pq} \cdot \cos(p\omega_a \pm q\omega_b)t, \quad . \quad . \quad . \quad (15)$$

and it has been shown (Ref. <sup>(3)</sup>) that the coefficients can be calculated by means of formulæ analogous to those equations (1) and (2). We have

$$C_{pq} = \sum_{k=0}^{k=K} S_{2k}(a, b) \times \delta^{2k} i_0, \quad . \quad . \quad . \quad (16a)$$

if  $p$  and  $q$  are both even,

$$C_{pq} = \sum_{k=0}^{k=K-1} S_{2k+1}(a, b) \times \mu \delta^{2k+1} i_0, \quad . \quad . \quad . \quad (16b)$$

if  $p$  or  $q$  is odd, or

$$C_{pq} = \sum_{k=0}^{k=K-1} S_{2k}(a, b) \times \mu^2 \delta^{2k} i_0, \quad . \quad . \quad . \quad (16c)$$

if both  $p$  and  $q$  are odd.

The functions  $S$  occurring in these formulæ represent certain combinations of the functions  $F$  we met in the previous section (equations (16) arise in fact from a multiplication of the appropriate expressions (1) or (2) for the appropriate coefficients  $C_p$  and  $C_q$  associated with the individual input frequencies).

The formulæ (16a) and (16b) are of exactly the same form as formulæ (1) and (2) and require therefore no further discussion. For the conversion of formula (16c) we have to evaluate the operator  $\mu^2$ . From (5) follows

$$\mu^2 = \frac{1}{4}[E^{+1} + 2 + E^{-1}]. \quad . \quad . \quad . \quad (17)$$

Hence

$$\mu^2 \delta^{2k} = \frac{1}{4}[E^{+1} + 2 + E^{-1}] \times \sum_{t=0}^{t=2k} (-1)^t C_t^{2k} E^{k-t}. \quad . \quad . \quad (18)$$

Substituting this expression into (16c) we get

$$C_{pq} = \frac{1}{4} \sum_{k=0}^{k=K-1} \sum_{t=0}^{t=2k} S_{2k}(a, b) \times (-1)^t \times [E^{k-t+1} + 2E^{k-t} + E^{k-t-1}], \quad (19)$$

i. e., we have here three double sums, one of the type of equation (8) and two of the type of equations (12). Hence the final result must be expressible in the form

$$C_{pq} = \sum_{w=-K}^{w=+K} i_w \times L_{pq}^w(a, b, K), \quad . \quad . \quad . \quad (20)$$

where now

$$L_{pq}^w(a, b, K) = \frac{1}{4} \sum_{k=0}^{k=K-1} S_{2k}(a, b) \times (-1)^{k-w} \times [-C_{k-w+1}^{2k} + 2C_{k-w}^{2k} - C_{k-w-1}^{2k}] \quad . \quad . \quad (21)$$

In Ref. <sup>(3)</sup> an example was dealt with where  $a=b=2$ ,  $K=4$ ,  $p=q=1$ . The Lagrangean coefficients required for this case were calculated by means of formula (19), using the results of the previous paper. The coefficients are given in Table II. For the sake of comparison the second

TABLE II.

$L^0$	=	-0.2777	-0.4444
$L^{+1}$	= $L^{-1}$	= -0.2888	-0.2222
$L^{+2}$	= $L^{-2}$	= +0.0444	+0.1111
$L^{+3}$	= $L^{-3}$	= +0.2888	+0.2222
$L^{+4}$	= $L^{-4}$	= +0.0944	+0.1111

column of this table gives the coefficients obtained by the use of the simpler method of Ref. <sup>(4)</sup>. The difference between the final result is, however, less than one would expect from looking at the two columns: the present method gives (in corroboration of the result of Ref. <sup>(3)</sup>)  $C=18.50mA$ , the method of Ref. <sup>(4)</sup> gives  $C=18.91mA$ . The reason for this small difference is of some interest, for, in fact, it can be shown that the two sets of coefficients will give the same result with any valve characteristic which is expressible by a polynomial of not higher order than the 5th.

It will be seen from Ref. <sup>(3)</sup> that the output  $C_1$  for the case where only one of the input frequencies is applied to the grid of the valve is given by the symbolic expression

$$C_1 = \left[ 2.0\mu\delta + \frac{2}{3}\mu\delta^3 - \frac{1}{60}\mu\delta^5 \right] \times i(0). \quad (22)$$

From this followed then the output component  $C_{1,1}$  for the case where the two input frequencies are simultaneously applied:

$$\begin{aligned} C_{1,1} &= \left[ 2.0\mu\delta + \frac{2}{3}\mu\delta^3 - \frac{1}{60}\mu\delta^5 \right]^2 \times i(0) \\ &= \left[ 4.0\mu^2\delta^2 + \frac{8}{3}\mu^2\delta^4 + \left( \frac{4}{9} - \frac{1}{15} \right) \mu^2\delta^6 + \dots \right] \times i(0). \quad (23) \end{aligned}$$

The functions  $S_2-S_6$  which we quoted above were the coefficients of the three terms lowest in  $\delta$  which arise from this expression.

Now, the simplified method of Ref. <sup>(4)</sup> amounts to the suppression of the last term in equation (22); hence we get in this case

$$C_{1,1} = \left[ 2.0\mu\delta + \frac{2}{3}\mu\delta^3 \right]^2 \times i(0) = \left[ 4.0\mu^2\delta^2 + \frac{8}{3}\mu^2\delta^4 + \frac{4}{9}\mu^2\delta^6 \right] \times i(0), \quad (24)$$

which yields the same values for  $S_2$  and  $S_4$ , but gives a slightly different value for  $S_6$ . ( $S_6=4/9$  instead of  $S_6=4/9-1/15$ .) It is this difference

which causes the large variation in the coefficients  $L$ . On this occasion we note the interesting conclusion that quite generally from equations (10), (13) or (21), by taking arbitrarily the values of the  $F$ - or  $S$ -functions up to the order  $m$  equal to zero, we arrive at sets of coefficients  $L$  which give zero result with any set of current ordinates provided these current ordinates all lie on a parabola of order not higher than  $m-1$ . This follows immediately from the fact that the equivalent formulae (1), (2) or (16) will give zero result, the first  $m$  product terms being zero on account of the assumption concerning the  $F$ - or  $S$ -functions, the rest of the terms being zero on account of the restriction of the order of the polynomial.

#### References.

- (1) Sets of such formulæ giving up to the 5th harmonic and taking into account up to seven equally spaced ordinates of the characteristic were given by D. C. Espley in Proceedings of the Institute of Radio Engineers, vol. xxi. pp. 1493-1446 (1933).
- (2) D. C. Espley, "Harmonic Analysis by the Method of Central Differences," Phil. Mag. vol. xxviii. pp. 338-352 (1939).
- (3) A. Bloch, "Calculation of Intermodulation Products by Means of a Difference Table," Journal of the Institution of Electrical Engineers, vol. xciii. Part III, p. 211 (1946).
- (4) A. Bloch, "Formulæ for the Direct Calculation of Intermodulation Products from the Valve Characteristic," 'Wireless Engineer,' vol. xxiii. p. 227 (1946).
- (5) Sheppard, Proceedings of the London Math. Soc. vol. xxxi. p. 459 (1899). See also : Whittaker & Robinson, 'Calculus of Observation,' Chapter I.

### LXXII. *Haidinger's Rings in Divergent Light.*

By G. N. RAMACHANDRAN (from the Department of Physics, Indian Institute of Science, Bangalore) \*.

[Received August 28, 1944.]

#### 1. *Introduction.*

THE "interference curves of equal inclination," first observed by Haidinger, which are exhibited by a transparent plate in the light transmitted or reflected by it, are of fundamental importance in optics, as they form the basis of the modern high resolving power interferometers, including especially the Lummer-Gehrcke plate and the Fabry-Perot etalon. In these applications, the plates employed have surfaces which are exactly plane and parallel, and the interferences are observed with an extended source of light, and a telescope focused for infinity. So much emphasis is usually laid upon these special conditions of observation that the impression has been created that they are essential for the observation of the Haidinger phenomenon. That, in reality, none of these conditions is indispensable was shown in a series of three papers<sup>(1), (2), (3)</sup>

\* Communicated by Sir C. V. Raman,



by Raman and Rajagopalan. In particular, they showed that an extended source of light is not necessary<sup>(1)</sup>, and that, with a point source, the interference fringes can be seen without an observing telescope directly on a viewing screen placed anywhere in the field of light reflected or transmitted by the plate. An explanation of these non-localized fringes was also developed, and it was shown that they owe their origin to the interference of an infinite series of beams diverging from the images formed by the successive reflections and refractions of the light at the two surfaces of the plate<sup>(2)</sup>.

It is the purpose of this paper to present a theory of the Haidinger interferences observed with a plane-parallel plate in light diverging from a point source on somewhat more general lines than those indicated by Raman and Rajagopalan. The treatment has the advantage that its validity is not restricted to the case of nearly normal incidence of the light on the plate. As is well-known, the multiple reflections at the surfaces, which must be considered even in the case of normal incidence, gain in intensity with increasing obliquity, and are responsible for the increased sharpness of the interferences observed in directions nearly parallel to the surface, a fact which is utilized in the functioning of the Lummer-Gehrcke interferometer.

## 2. Approximate Theory.

We shall consider the case of a parallel plate of thickness  $t$  and of refractive index  $\mu$  placed at a distance  $R_1$  from a point source of light  $S$ .

Fig. 1 (a).

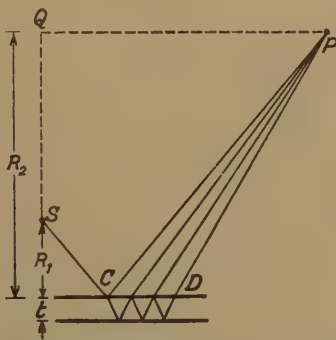
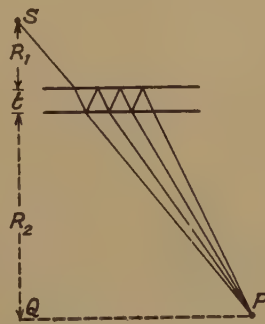


Fig. 1 (b).



The normal distance of any point of observation  $P$  from the plate may be denoted by  $R_2$ . It is clear that the illumination at  $P$  arises from the superposition of an infinite number of rays—viz., one directly reflected from the front surface, one which suffers two internal reflections, and so on. It is obvious that the actual path of every one of these rays will be different from those of the others.

We shall use the Fermat principle—viz., that the actual path in any case is a minimum—to make an approximate calculation of the path differences between the various rays. We assume the paths of the various rays to be as shown in fig. 1 (a), these being slightly different

from the actual paths; but by Fermat's principle, it follows that the error thus introduced would be small. We suppose that the  $n$ th ray follows the same path as the directly reflected ray SCP from the source up to the plate, and then follows a path within the medium which this ray would take up to the point of emergence D (suffering  $2n$  internal reflections). From this point, we imagine the ray to reach P along the path DP. We can now calculate the path difference between the directly reflected ray and this  $n$ th ray.

Let  $\theta$  be the angle made with the normal to the plate by the directly reflected ray, and let  $\phi$  be the corresponding angle of refraction. It is then easily seen that, approximately,

$$CP - DP = CD \sin \theta - \frac{CD^2 \cos^2 \theta}{2CP} = 2nt \tan \phi \sin \theta - \frac{2n^2 t^2 \tan^2 \phi \cos^2 \theta}{CP}. \quad (1)$$

The path difference between the directly reflected ray and the  $n$ th ray is therefore

$$\delta_n = 2\mu nt \cos \phi + 2n^2 t^2 \tan^2 \phi \cos^2 \theta / CP. \quad (2)$$

Hence, if we restrict ourselves to the particular case when the optical path from the plate to the point of observation is large compared with the thickness of the plate,  $\delta_n$  can be put equal to  $2\mu nt \cos \phi^*$ , the error so introduced being small. The path retardations of the successive rays thus form an arithmetic progression. Also, if we denote by  $b, d, c, f$ , the reflection coefficients of the surface in air and in the medium, and the transmission coefficients from air to the medium, and *vice versa* respectively, their amplitudes are proportional to  $cdf, cd^3f, \dots$ . The total amplitude at P is thus proportional to

$$b + cdf e^{-i\epsilon} + cd^3f e^{-2i\epsilon} + \dots,$$

where  $\epsilon = (2\pi/\lambda)(2\mu t \cos \phi)$ . Therefore the intensity at P comes out to be proportional to

$$\frac{4b^2 \sin^2(\epsilon/2)}{1 - 2b^2 \cos \epsilon + b^4} = \frac{4b^2}{(1 - b^2)^2 / \sin^2(\epsilon/2) + 4b^2}, \quad (3)$$

using the well-known relations  $b = -d$ , and  $cf = 1 - b^2$ .

In an exactly similar manner, if  $\theta$  and  $\phi$  are the angles made with the normal in air and in the medium respectively by the ray reaching P directly after transmission through the plate (fig. 1(b)), and if  $b$  is the reflection coefficient corresponding to these angles, then the intensity at P is, to a first approximation, proportional to

$$(1 - b^2)^2 / [(1 - b^2)^2 - 4b^2 \cos^2(\epsilon/2)]. \quad (4)$$

The above expressions (3) and (4) are identical with those for the variation of intensity with angle for the Haidinger's rings formed at infinity. The reflected beam has minima, and the transmitted beam maxima, corresponding to  $\epsilon = 2m\pi$ , and *vice versa* for  $\epsilon = (2m + 1)\pi$ . The

---

\* Actually, as will be seen in the next section, the condition for putting  $\delta_n = 2\mu nt \cos \phi$  is that the total optical path from the source to the point of observation is small compared with the thickness of the plate.

angular radii of the rings are, to a first approximation, the same as those formed at infinity, while the linear radius increases in direct proportion to the total distance from the source.

It may be seen from equation (2) that the error in  $\delta_n$  vanishes identically for grazing incidence for which  $\cos^2\theta=0$ , and also for normal incidence, since  $\tan^2\phi=0$  in this case. Thus, the non-localized fringes formed with a point source and the Lummer-Gehrcke plate will exhibit the same features as at infinity, provided the plane of observation is sufficiently far from the source. In the Fabry-Perot interferometer, the incidence is nearly normal, and the surfaces are silvered to increase the reflecting power. It is obvious from the above theory that in this case also one should expect sharp non-localized fringes at a sufficient distance from the source\*.

If one employs an extended source, then each point in it will give rise to a non-localized interference pattern, which will all be superposed. The result is that the pattern will not be visible at finite distances. One has to go to infinity to avoid the blurring due to the superposition. The smaller the dimensions of the source the nearer can the observation screen be brought to the source without the fringes suffering appreciable loss of visibility. The condition to be satisfied is that the angular width of the source as seen from the point of observation must be small compared with the angular separation of the fringes. In fact, this idea of the superposition of the interference patterns due to individual point sources was given by Wood<sup>(4)</sup>, although the fact that a point source gives non-localized fringes does not appear to have been noticed prior to the work of Raman and Rajagopalan. These considerations also fully substantiate the statement made by Sir C. V. Raman<sup>(6)</sup> that the non-localized fringes obtained with a point source are essentially of the same nature as the Haidinger interferences. Indeed, it is clear that the use of an extended source in the observation of the Haidinger phenomena serves only to increase the intensity of illumination, while, *per contra*, it restricts the plane of observation to the plane at infinity, or the focal plane of a viewing telescope.

### 3. More Exact Considerations.

In the preceding section, it was assumed that the various rays contributing to the intensity at P could be considered as following the same path as the directly reflected or transmitted ray up to the final point of emergence from the plate. In this section, our purpose is to calculate

---

\* The special case of the Fabry-Perot etalon, *i. e.* a plane-parallel plate of refractive index unity, has been discussed by Tolansky<sup>(5)</sup> on the basis of the formation of multiple virtual images as suggested earlier by Raman and Rajagopalan<sup>(2)</sup>. It is obvious that this procedure cannot be adopted if the refractive index is different from unity, for the geometric position of the images varies with the angle of incidence. It may also be remarked that the statement made by Tolansky that, in an unsilvered plate, only two effective beams interfere is not correct. Actually, one must consider a whole series of reflected and transmitted beams.



the path retardations more rigorously, and to show that the deviation from the arithmetic progression value ( $2\mu nt \cos \phi$ ) is small, provided the total optical path from the source to the point of observation is large compared with the thickness of the plate. We shall consider only the cases of nearly grazing and nearly normal incidences.

Let  $\theta_n$  and  $\phi_n$  be the angles made by the  $n$ th ray with the normal to the plate in air and in the medium. Considering the reflected system obtained with the Lummer-Gehrcke plate, the angle of incidence is nearly  $\pi/2$ , and we may denote  $\pi/2 - \theta$  by  $\alpha$ , and  $\pi/2 - \theta_n$  by  $\alpha_n$ , where  $\alpha$  and  $\alpha_n$  are small. If  $R$  represents the distance of the screen from the source measured parallel to the plate (PQ in fig. 1 (a)), then it can be shown that

$$\alpha_n - \alpha = \alpha \left[ \frac{2nt}{R\sqrt{\mu^2 - 1}} + \frac{4n^2 t^2}{R^2(\mu^2 - 1)} \right]. \quad (5)$$

The path difference between the directly reflected ray and the  $n$ th ray comes out (to the second degree of approximation) to be equal to

$$2\mu nt \cos \phi + 2n^2 t^2 \alpha^2 / R(\mu^2 - 1). \quad (6)$$

Hence the deviation of the path of the  $n$ th ray from the value  $2\mu nt \cos \phi$  is

$$\Delta \delta_n = 2n^2 t^2 \alpha^2 / R(\mu^2 - 1). \quad (7)$$

For the first fringe (the one closest to the plate),

$$\alpha \simeq \sqrt{(\mu^2 - 1)^{1/2}} \lambda / t, \quad \text{and} \quad \Delta \delta_n \simeq 2n^2 t \lambda / R \sqrt{\mu^2 - 1}. \quad (8)$$

This is of the order of  $10^{-3}$  of a wave-length for  $n=1$ , with a plate of thickness 2 mm. and refractive index 1.5, the distance  $R$  being 1 metre. For the rays which suffer more internal reflections, the deviation is larger; but their intensity is much less. Thus the sharpness of the first few fringes is not appreciably affected by having the plane of observation at a finite distance from the source. For the outer fringes, the value of  $\alpha$  is larger, and  $\Delta \delta_n$  is not negligible, so that the sharpness of the outer rings is not so good as at the centre. Also, as the plane of observation is brought nearer and nearer to the source, fringes corresponding to smaller glancing angles also begin to be progressively affected. It may also be pointed out that the rings are sharper, the greater the refractive index, the smaller the thickness, and the farther the point of observation is from the source.

Precisely the same phenomena occur with the transmitted system also. If we denote by  $\alpha$  and  $\alpha_n$  the glancing angles of incidence for the directly transmitted ray, and the ray that suffers  $2n$  internal reflections, respectively,

$$\alpha_n - \alpha = \alpha \left[ \frac{2nt}{L\sqrt{\mu^2 - 1}} + \frac{4n^2 t^2}{L^2(\mu^2 - 1)} \right], \quad \text{where } L = R - \frac{t}{\sqrt{\mu^2 - 1}}, \quad (9)$$

and

$$\Delta \delta_n = 2n^2 t^2 \alpha^2 / L(\mu^2 - 1), \quad (10)$$

expressions very similar to (5) and (7).



For the case of nearly normal incidence, viz. one analogous to the Fabry-Perot etalon,

$$\theta_n - \theta = \theta \left[ -\frac{2nt}{\mu(R_1 + R_2)} + \frac{4n^2t^2}{\mu^2(R_1 + R_2)^2} \right], \quad \dots \quad (11)$$

and

$$\Delta\delta_n = 2n^2t^2\phi^2/(R_1 + R_2), \quad \dots \quad (12)$$

expressions similar to those obtained with the Lummer-Gehrcke plate. The change in path is very small for the rings close to the centre of the field and all the other statements made about the Lummer-Gehrcke plate also hold for this case. For the transmitted system,  $\theta_n - \theta$  and  $\Delta\delta_n$  are the same as (11) and (12), except that  $\mu(R_1 + R_2)$  is to be replaced by  $\mu(R_1 + R_2) + t$ .

#### 4. Haidinger's Rings in Birefringent Plates.

It is obvious that if one employs a plate of mica, quartz or any such birefringent crystal for the observation of the Haidinger phenomenon, then two systems of rings would be formed. This was first remarked by the late Lord Rayleigh<sup>(7)</sup>. Chinmayanandam<sup>(8)</sup> studied the phenomenon in some detail, and showed that the two sets of rings are approximately elliptical in shape, the superposition of which results in a fluctuation of the visibility of the rings seen in the field of view. In the case of mica, which is a biaxial crystal of which the faces are nearly perpendicular to the bisectrix between the optic axes, Chinmayanandam also showed that the form of the lines of minimum visibility is identical with the "isochromatic lines" in polarized light for a plate of twice the thickness. Ghosh<sup>(9)</sup> investigated the Haidinger rings in plates of quartz cut in different directions. By silvering the surfaces of the plate, he was able to observe the two systems of rings well resolved. Later, Schaefer and his collaborators<sup>(10)</sup> studied the phenomena at oblique incidence and also considered the theory of these interferences in some detail. Recently, Pfund<sup>(11)</sup> has investigated the Haidinger rings in silvered plates of mica and has obtained some very striking photographs showing the variations in the visibility of the rings, and also the identity of the "moiré" patterns with the polarization bands of plates of double the thickness.

It is obvious that all these phenomena should be capable of observation with a point source of light, when non-localised fringes would be formed under the same conditions as they are formed with an isotropic plate.

My best thanks are due to Prof. Sir C. V. Raman for the many helpful suggestions which he gave me during the course of the investigation.

#### Summary.

The theory of the Haidinger rings with a plane-parallel plate and a point source is worked out in this paper by considering the path retardation of the various rays that contribute to the intensity at the point of observation. It is shown that the interference fringes are non-localized, and that they should be visible on a screen kept at any point on either side of the plate. Also, if the total path from the source to the point of

observation is large compared with the thickness of the plate, then the fringes would be as clearly defined as at infinity, the path retardations of the various rays being the same as for infinity to a first approximation. The deviations occurring in the path, which are small under the above circumstances, are particularly negligible for angles of incidence near about  $90^\circ$  and  $0^\circ$ , vanishing completely for exactly grazing and exactly normal incidences. Thus, with the Lummer-Gehrcke and the Fabry-Perot interferometers, and a point source, sharp fringes must be visible at a sufficient distance from the source. The extension of the results of the theory to the case of birefringent plates is also pointed out.

### Bibliography.

- (1) Sir C. V. Raman and V. S. Rajagopalan, Journ. Opt. Soc. Am. xxix. p. 413 (1939).
- (2) Sir C. V. Raman and V. S. Rajagopalan, Proc. Ind. Acad. Sci. A, x. p. 317 (1939).
- (3) Sir C. V. Raman and V. S. Rajagopalan, Phil. Mag. xxix. p. 508 (1940).
- (4) R. W. Wood, 'Physical Optics,' 3rd edition, p. 208.
- (5) S. Tolansky, Phil. Mag. xxxiv. p. 555 (1943); xxxv. p. 211 (1944).
- (6) Sir C. V. Raman, Phil. Mag. xxxv. p. 210 (1944).
- (7) Lord Rayleigh, Phil. Mag. xii. p. 489 (1906).
- (8) T. K. Chinmayanandam, Proc. Roy. Soc. A, xcv. p. 176 (1918).
- (9) P. N. Ghosh, Proc. Ind. Assoc. Cult. Sci. vii. p. 53 (1922).
- (10) C. Schaefer *et al.*, *Zeits. f. Phys.* xiv. p. 253 (1923); xvii. p. 155 (1923); *Zeits. f. Tech. Phys.* iii. p. 98 (1926).
- (11) A. H. Pfund, Journ. Opt. Soc. Am. xxxii. p. 383 (1942).

## LXXIII. *A Discussion of the Pressure-tendencies associated with Gradient and Horizontal Geostrophic Flow. A Formula for the Variation with Height of the Vertical Velocity.*

By A. G. MATTHEWMAN, M.A.\*

[Received April 4, 1946.]

### 1. Introduction.

THE pressure-tendency at any place is customarily expressed in terms of an integration over the column of air vertically above the place. This form of expression arises from the use of the hydrostatic equation.

In the present paper, no use will be made of the hydrostatic equation. The pressure-tendency will be derived in terms of the fields of motion, of temperature, and of humidity in a neighbourhood of the place.

### 2. Assumptions.

The following conditions will be assumed to hold :—

- (i) The air is unsaturated.
- (ii) In its motion it satisfies the dry-adiabatic relations.

\* Communicated by Sir Nelson Johnson, K.C.B., D.Sc.

- (iii) Throughout a *horizontal* neighbourhood of the place at which the pressure-tendency is required, the velocity of the air is gradient.

Horizontal geostrophic flow is considered as a special case of gradient flow.

In the last part of the last section (10), the condition (ii) is withdrawn, the formula for the variation with height of the vertical velocity thus holding under conditions (i) and (iii).

### 3. A Formula for Pressure-tendency in Terms of Divergence at the Place.

The dry-adiabatic condition may be written

$$\frac{1}{\gamma p} \frac{dp}{dt} = \frac{1}{\rho} \frac{d\rho}{dt}, \quad \dots \dots \dots (1)$$

where  $\gamma$ =ratio of the specific heat at constant pressure to the specific heat at constant volume.

If  $\mathbf{G}$  denotes the gradient wind vector, since  $\mathbf{G}$  is tangential to the horizontal isobar at the place,

$$\mathbf{G} \cdot \nabla p = 0.$$

Therefore where the velocity is gradient

$$\frac{dp}{dt} = \frac{\partial p}{\partial t} + \mathbf{G} \cdot \nabla p = \frac{\partial p}{\partial t}. \quad \dots \dots \dots (2)$$

Hence, if the dry-adiabatic condition holds, and the velocity at the place is gradient, equations (1) and (2) give

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = \frac{1}{\rho} \frac{d\rho}{dt}. \quad \dots \dots \dots (3)$$

The equation of continuity can be written

$$\frac{1}{\rho} \frac{d\rho}{dt} = -\nabla \mathbf{V}, \quad \dots \dots \dots (4)$$

where  $\mathbf{V}$  is the velocity vector.

But 
$$\nabla \mathbf{V} = \nabla_h \mathbf{V} + \frac{\partial w}{\partial z},$$

where  $\nabla_h$  denotes the horizontal divergence,  $z$  is measured vertically upwards, and  $w$  is the vertical component of velocity.

If the velocity throughout a horizontal neighbourhood is gradient

$$\nabla_h \mathbf{V} = \nabla_h \mathbf{G} = \nabla \mathbf{G},$$

therefore

$$\nabla \mathbf{V} = \nabla \mathbf{G} + \frac{\partial w}{\partial z}. \quad \dots \dots \dots (5)$$

It follows, under the assumed conditions specified in section 2, from equations (3), (4) and (5), that

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = -\nabla \mathbf{G} - \frac{\partial w}{\partial z}. \quad \dots \dots \dots (6)$$

It is convenient to write

$$\nabla \mathbf{G} \equiv \frac{\nabla(\rho \mathbf{G})}{\rho} - \frac{\mathbf{G} \cdot \nabla \rho}{\rho},$$

so that equation (6) becomes

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = -\frac{\nabla(\rho \mathbf{G})}{\rho} + \frac{\mathbf{G} \cdot \nabla \rho}{\rho} - \frac{\partial w}{\partial z} \quad (7)$$

The terms  $\frac{\nabla(\rho \mathbf{G})}{\rho}$  and  $\frac{\mathbf{G} \cdot \nabla \rho}{\rho}$  on the right-hand side of equation (7) will now be evaluated in turn.

#### 4. Evaluation of $\frac{\nabla(\rho \mathbf{G})}{\rho}$ .

Whatever the actual velocity of the air (provided it is not purely vertical) the gradient-wind vector  $\mathbf{G}$  may be supposed to exist and to satisfy the defining equation

$$(\mathbf{G}\kappa + \lambda)\mathbf{G} = \lambda\mathbf{J}, \quad (8)$$

where  $G$  is the magnitude of  $\mathbf{G}$ , that is, the gradient-wind speed;  $\kappa$  is the curvature of the vertical projection on to the horizontal plane of the trajectory of the element of air, positive for cyclonic, negative for anti-cyclonic curvature;  $\lambda = 2\omega \sin \phi$ ,  $\omega$  = rate of rotation of the earth about its axis,  $\phi$  denotes latitude; and  $\mathbf{J}$  is the geostrophic-wind vector.

Multiplying equation (8) by the density  $\rho$ , and taking the divergence of each side:

$$\nabla\{(\mathbf{G}\kappa + \lambda)\rho\mathbf{G}\} = \nabla(\lambda\rho\mathbf{J}). \quad (9)$$

Horizontal geostrophic motion is defined as horizontal frictionless motion unaccelerated relative to the earth, thus

$$\lambda\rho\mathbf{J} = [\mathbf{I} \wedge \nabla p],$$

where  $\mathbf{I}$  denotes unit vertical vector.

The same equation serves to define  $\mathbf{J}$  at any point in the atmosphere whatever the actual velocity of the air may be.

It may be thought sufficient, in order to evaluate  $\nabla(\lambda\rho\mathbf{J})$ , to take rectangular Cartesian co-ordinates with the origin at the point considered, and the axis of  $z$  vertical, and to write

$$\lambda\rho\mathbf{J} = [\mathbf{I} \wedge \nabla p] = \left(-\frac{\partial p}{\partial y}, \frac{\partial p}{\partial x}, 0\right).$$

But it will be noticed that

$$[\mathbf{I} \wedge \nabla p] \neq \left(-\frac{\partial p}{\partial y}, \frac{\partial p}{\partial x}, 0\right)$$

except at points on the axis of  $z$ , for  $[\mathbf{I} \wedge \nabla p]$  is everywhere horizontal, while  $\left(-\frac{\partial p}{\partial y}, \frac{\partial p}{\partial x}, 0\right)$  is everywhere parallel to the co-ordinate plane of  $x, y$ , which is horizontal only at the origin. Although the difference between the two vectors at any point is obviously small and tends to zero as the point tends towards the axis of  $z$ , it is not obvious what magnitude of error may be involved in inferring that because

$$\frac{\partial}{\partial x} \left(-\frac{\partial p}{\partial y}\right) + \frac{\partial}{\partial y} \left(\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial x}(0) = 0,$$

therefore

$$\nabla(\lambda\rho\mathbf{J}) = 0.$$



To avoid difficulties of this kind, consider the vectorial definition of the unit vertical vector  $\mathbf{I}$ . If  $u$  denotes the combined gravitational and centrifugal potential at the point considered,

$$\nabla u = -g\mathbf{I},$$

where  $g$  denotes the value of apparent gravity.

$$\text{Thus} \quad \lambda\rho\mathbf{J} = [\mathbf{I} \wedge \nabla p] = \frac{1}{g} [\nabla p \wedge \nabla u].$$

$$\text{Hence} \quad \nabla(\lambda\rho\mathbf{J}) = \nabla \left\{ \frac{1}{g} [\nabla p \wedge \nabla u] \right\}.$$

But  $\nabla[\nabla p \wedge \nabla u] \equiv 0$ , therefore

$$\begin{aligned} \nabla(\lambda\rho\mathbf{J}) &= -\frac{1}{g^2} \nabla g \cdot [\nabla p \wedge \nabla u] \\ &= \frac{\nabla g}{g} \cdot [\nabla p \wedge \mathbf{I}]. \end{aligned}$$

$$\text{Thus} \quad \nabla(\lambda\rho\mathbf{J}) = -\frac{\nabla g}{g} \cdot (\lambda\rho\mathbf{J}).$$

Now  $\mathbf{J}$  is a horizontal vector, and in the horizontal apparent gravity varies only with latitude, its value in latitude  $\phi$  being related to its value in latitude  $45^\circ$  by the equation

$$g_\phi = g_{45}(1 - 0.0026 \cos 2\phi).$$

$$\begin{aligned} \text{Thus} \quad \frac{\nabla g}{g} &= (0.0052 \sin 2\phi) \nabla \phi \\ &= (0.0052 \sin 2\phi) \frac{\partial \phi}{\partial y} \mathbf{Y}, \end{aligned}$$

where  $\mathbf{Y}$  denotes unit horizontal vector directed from south to north.

$$\text{But} \quad \frac{\partial \phi}{\partial y} = \frac{1}{R^*}$$

where  $R^*$  denotes the radius of curvature of a meridian section of the level surface  $u = \text{constant}$  at the point.

With sufficient approximation

$$R^* = R_E,$$

the average radius of the earth, about 4000 miles. Thus

$$\frac{\nabla g}{g} = \frac{0.0052 \sin 2\phi}{R_E} \mathbf{Y},$$

$$\begin{aligned} \text{and} \quad \frac{\nabla g}{g} \cdot (\lambda\rho\mathbf{J}) &= \frac{0.0052 \sin 2\phi}{R_E} \mathbf{Y} \cdot (\lambda\rho\mathbf{J}) \\ &= \frac{0.0052 \sin 2\phi}{R_E} \lambda\rho J_{S \rightarrow N}, \end{aligned}$$

where  $J_{S \rightarrow N}$  denotes the south-north component of the geostrophic wind at the point.

Thus 
$$\nabla(\lambda\rho\mathbf{J})=-\frac{0.0052\sin 2\phi}{R_E}\lambda\rho\mathbf{J}_{S\rightarrow N}.$$

In what follows, a magnitude of the order of  $\frac{\lambda\rho\mathbf{J}_{S\rightarrow N}}{R_E}$  in the divergence will prove important, but 1/200 of this magnitude may be neglected. To this order of approximation,

$$\nabla(\lambda\rho\mathbf{J})=0.$$

From equation (9) it follows that, to the same order of approximation,

$$\nabla\{(G\kappa+\lambda)\rho\mathbf{G}\}=0.$$

Therefore 
$$(G\kappa+\lambda)\nabla(\rho\mathbf{G})+\rho\mathbf{G}.\nabla(G\kappa+\lambda)=0.$$

Thus 
$$\frac{\nabla(\rho\mathbf{G})}{\rho}=-\frac{\mathbf{G}.\nabla(G\kappa+\lambda)}{G\kappa+\lambda} \dots \dots \dots (10)$$

Since 
$$\nabla(G\kappa+\lambda)=\kappa\nabla G+G\nabla\kappa+\nabla\lambda,$$

equation (10) becomes

$$\frac{\nabla(\rho\mathbf{G})}{\rho}=-\frac{1}{G\kappa+\lambda}\{\kappa\mathbf{G}.\nabla G+G\mathbf{G}.\nabla\kappa+\mathbf{G}.\nabla\lambda\}. \dots \dots (11)$$

Consider first  $\mathbf{G}.\nabla\lambda$ .

Since 
$$\begin{aligned} \lambda &= 2\omega \sin \phi \\ \nabla\lambda &= 2\omega \cos \phi \nabla\phi \\ &= \frac{2\omega \cos \phi}{R_E} \mathbf{Y} \end{aligned}$$

where, with sufficient approximation, the radius of curvature of the meridian section of the level surface at the point has been replaced by the average radius of the earth  $R_E$ .

Then 
$$\mathbf{G}.\nabla\lambda=\frac{2\omega \cos \phi}{R_E}\mathbf{G}.\mathbf{Y}=\frac{2\omega \cos \phi}{R_E}G_{S\rightarrow N}, \dots \dots \dots (12)$$

where  $G_{S\rightarrow N}$  denotes the south-north component of the gradient wind-vector.

In the other terms within the curly bracket on the right-hand side of equation (11), let  $\frac{\partial}{\partial s}$  denote differentiation along the isobar in the direction of  $\mathbf{G}$ , so that

$$\mathbf{G}.\nabla\equiv G\frac{\partial}{\partial s}. \dots \dots \dots (13)$$

Equation (11) can thus be written, in virtue of (12) and (13),

$$\frac{\nabla(\rho\mathbf{G})}{\rho}=-\frac{1}{G\kappa+\lambda}\left\{\kappa G\frac{\partial G}{\partial s}+G^2\frac{\partial \kappa}{\partial s}+\frac{2\omega \cos \phi}{R_E}G_{S\rightarrow N}\right\}. \dots \dots (14)$$

### 5. Evaluation of $\frac{\mathbf{G} \cdot \nabla \rho}{\rho}$ .

Since the air is, by hypothesis, unsaturated, at ordinary temperatures the equation of state,

$$p = R \left\{ \frac{1+x/\epsilon}{1+x} \right\} \rho T, \quad . . . . . (15)$$

is satisfied to a high degree of approximation.  $x$  is the humidity mixing ratio,  $\epsilon$  is the specific gravity of water-vapour ( $=0.622$  or  $5/8$ , approximately), and  $R$  is the constant appropriate to dry air.

Since  $x$  is of the order of  $1/100$  (gram. of water-vapour per gm. of dry air), equation (15) may be written with sufficient approximation,

$$p = R \left\{ 1 + x \left( \frac{1}{\epsilon} - 1 \right) \right\} \rho T,$$

or

$$p = R \left( 1 + \frac{3}{5} x \right) \rho T. \quad . . . . . (16)$$

Taking first the logarithm and then the gradient of each side of equation (16),

$$\frac{\nabla p}{p} = \frac{3}{5} \nabla x + \frac{\nabla \rho}{\rho} + \frac{\nabla T}{T},$$

with sufficient approximation.

Multiplying this last equation scalarly by  $\mathbf{G}$ , since  $\mathbf{G} \cdot \nabla p = 0$ ,

$$\frac{\mathbf{G} \cdot \nabla \rho}{\rho} = -\frac{3}{5} \mathbf{G} \cdot \nabla x - \frac{\mathbf{G} \cdot \nabla T}{T},$$

or

$$\frac{\mathbf{G} \cdot \nabla \rho}{\rho} = -\frac{3}{5} G \frac{\partial x}{\partial s} - \frac{G}{T} \frac{\partial T}{\partial s}. \quad . . . . . (17)$$

### 6. The Pressure-tendency Associated with Dry-adiabatic Gradient Flow.

Assuming the conditions specified in section 2, it follows from equations (7), (14) and (17) derived above, that

$$\begin{aligned} \frac{1}{\gamma p} \frac{\partial p}{\partial t} = \frac{1}{G\kappa + \lambda} \left\{ \kappa G \frac{\partial G}{\partial s} + G^2 \frac{\partial \kappa}{\partial s} + \frac{2\omega \cos \phi}{R_E} G_{S \rightarrow N} \right\} \\ - \frac{3}{5} G \frac{\partial x}{\partial s} - \frac{G}{T} \frac{\partial T}{\partial s} - \frac{\partial w}{\partial z}. \quad . . . (18) \end{aligned}$$

### 7. The Pressure-tendency Associated with Dry-adiabatic Gradient Flow Geostrophic at the Place.

From the defining equation of the gradient-wind vector,

$$(G\kappa + \lambda)\mathbf{G} = \lambda\mathbf{J},$$

it follows that when  $\kappa=0$ ,  $\mathbf{G}=\mathbf{J}$ , and conversely (excluding the trivial  $\mathbf{G}=0$ ).

Hence the formula for the pressure-tendency in this case is given by substituting in equation (18)  $\kappa=0$ ,  $G=J$  (the geostrophic wind-speed)

and  $G_{S \rightarrow N} = J_{S \rightarrow N}$  (the magnitude of the south-north geostrophic component.) Thus

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = \frac{1}{\lambda} \left\{ J^2 \frac{\partial \kappa}{\partial s} + \frac{2\omega \cos \phi}{R_E} J_{S \rightarrow N} \right\} - \frac{3}{5} J \frac{\partial x}{\partial s} - \frac{J}{T} \frac{\partial T}{\partial s} - \frac{\partial w}{\partial z}. \quad (19)$$

Since  $\lambda = 2\omega \sin \phi$ , the term

$$\frac{1}{\lambda} \frac{2\omega \cos \phi}{R_E} J_{S \rightarrow N} = \frac{\cot \phi}{R_E} J_{S \rightarrow N},$$

and equation (19) becomes

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = \frac{1}{\lambda} J^2 \frac{\partial \kappa}{\partial s} + \frac{\cot \phi}{R_E} J_{S \rightarrow N} - \frac{3}{5} J \frac{\partial x}{\partial s} - \frac{J}{T} \frac{\partial T}{\partial s} - \frac{\partial w}{\partial z}. \quad (20)$$

8. *The Pressure-tendency Associated with Dry-adiabatic Gradient Flow, Geostrophic along the Horizontal Isobar in a Neighbourhood of the Place.*

In this case,  $\frac{\partial \kappa}{\partial s} = 0$  in equation (20). Thus

$$\frac{1}{\gamma p} \frac{\partial p}{\partial t} = \frac{\cot \phi}{R_E} J_{S \rightarrow N} - \frac{3}{5} J \frac{\partial x}{\partial s} - \frac{J}{T} \frac{\partial T}{\partial s} - \frac{\partial w}{\partial z}. \quad (21)$$

9. *Numerical Examples Illustrating Equations (21) and (20).*

(i) Unsaturated air, under dry adiabatic conditions; horizontal flow throughout a neighbourhood of the place, gradient flow throughout a horizontal neighbourhood of the place, geostrophic along the horizontal isobar in a neighbourhood of the place.

By hypothesis  $\frac{\partial w}{\partial z} = 0$ . Taking the values

$$p = 950 \text{ mb.}, \quad \phi = 45^\circ, \quad R_E = 4,000 \text{ miles}, \quad J_{S \rightarrow N} = 30 \text{ m.p.h.},$$

$$J = 40 \text{ m.p.h.}, \quad T = 280^\circ \text{ A.}, \quad \frac{\partial T}{\partial s} = -1^\circ \text{ A. per 100 miles}, \quad \frac{\partial x}{\partial s} = 0,$$

equation (21) gives

$$\frac{\partial p}{\partial t} = 12 \text{ mb. per hour.} \quad (22)$$

A pressure-tendency so large probably never occurs.

The term  $\frac{\cot \phi}{R_E} J_{S \rightarrow N}$  on the right-hand side of equation (21) contributes about 10 mb. per hour towards the pressure-tendency given by (22).

Therefore, taking other and ordinary values of  $\frac{\partial T}{\partial s}$ ,  $\frac{\partial x}{\partial s}$  and  $J$  only slightly modifies this result.

It follows, then, that in middle and low latitudes, with a south-north component of the order of 30 m.p.h., the conditions specified at the beginning of this example are in general inconsistent with an assumption of zero pressure-tendency. In particular, horizontal geostrophic flow throughout a neighbourhood of the place is in general inconsistent with an assumption of zero pressure tendency. Even pressure-tendencies of the magnitude normally observed are inadmissible unless one or more of the conditions specified at the beginning of this example be relaxed.



It will next be shown that it may be sufficient to relax the condition that the flow throughout a neighbourhood of the place is horizontal, that is, to postulate vertical motion above and below the place, while retaining gradient, and therefore horizontal, motion throughout a horizontal neighbourhood of the place; and/or the condition that the wind is geostrophic along the horizontal isobar in a neighbourhood of the place, that is, to postulate geostrophic motion only at one point of the isobar, namely the place itself.

The two examples following will also show the necessary extent of these relaxations.

(ii) Unsaturated air, under dry-adiabatic conditions; gradient flow throughout a horizontal neighbourhood of the place, geostrophic along the horizontal isobar in a neighbourhood of the place.

Taking the same values as before, suppose in addition that  $\frac{\partial p}{\partial t} = 0$ .

Then equation (21) gives

$$\begin{aligned}\frac{\partial w}{\partial z} &= 0.009 \text{ m.p.h. per vertical mile} \\ &= 0.25 \text{ cm. per sec. per vertical km.}\end{aligned}$$

Taking, instead of  $\frac{\partial p}{\partial t} = 0$ ,  $\frac{\partial p}{\partial t} = 2$  mb. per hour,  $\frac{\partial w}{\partial z} = 0.21$  cm. per sec. per vertical km.

If  $\frac{\partial p}{\partial t} = -2$  mb. per hour,  $\frac{\partial w}{\partial z} = 0.29$  cm. per sec. per vertical km.

At the place, where the wind is, by hypothesis, geostrophic,  $w = 0$ . A "normal" pressure-tendency is thus consistent with the conditions specified at the beginning of this example, with a south-north component of the order of 30 m.p.h. provided there is a gradient of vertical velocity of the order of magnitude of 0.25 cm. per sec. per vertical km.

(iii) Unsaturated air, under dry-adiabatic conditions; horizontal flow throughout a neighbourhood of the place, gradient flow throughout horizontal neighbourhood of the place, geostrophic at the place itself.

By hypothesis  $\frac{\partial w}{\partial z} = 0$ .

Taking the same values as in example (i), suppose in addition that

$$\frac{\partial p}{\partial t} = 0.$$

Then equation (20) gives

$$\frac{\partial \kappa}{\partial s} = -\frac{1}{480,000} \text{ miles}^{-2}. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Taking instead of  $\frac{\partial p}{\partial t} = 0$ ,  $\frac{\partial p}{\partial t} = 2$  mb. per hour,

$$\frac{\partial \kappa}{\partial s} = -\frac{1}{580,000} \text{ miles}^{-2}, \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Or, if  $\frac{\partial p}{\partial t} = -2$  mb. per hour,

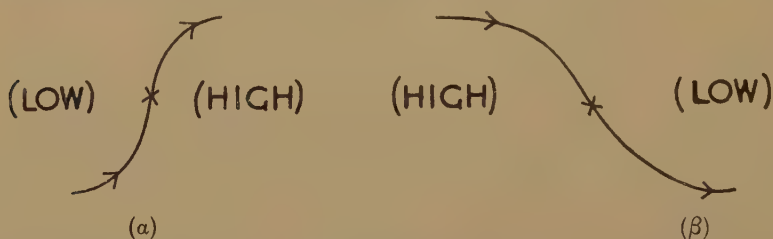
$$\frac{\partial \kappa}{\partial s} = -\frac{1}{410,000} \text{ miles}^{-2}. \quad . \quad . \quad . \quad . \quad . \quad (25)$$

At the place, where the wind is, by hypothesis, geostrophic,  $\kappa=0$ .

“Normal” pressure-tendencies are thus consistent with the conditions specified at the beginning of this example, provided that there is a change of curvature of the trajectory along the isobar, the rate of change in the direction of the geostrophic wind being of the order of

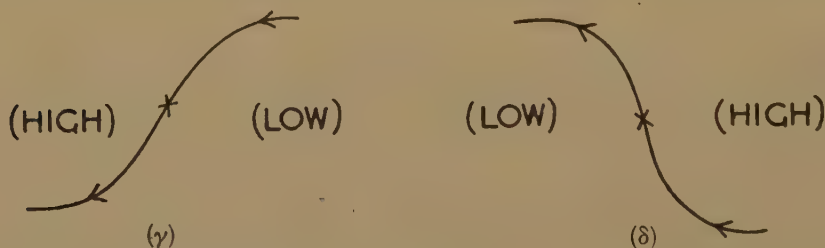
$$-\frac{1}{5000} \text{ miles}^{-1} \text{ per 100 miles.}$$

If the motion may be supposed steady, the lines of motion and the paths of the air particles coincide as stream-lines. From equation (20) it follows that when the conditions of this example are fulfilled the shape of a stream-line in the neighbourhood of a point of inflexion must be as indicated below in diagram ( $\alpha$ ) for the south-north case.



The north-south case is easily shown to be as in ( $\beta$ ) above. These shapes of stream-lines near their point of inflexion are typical of the shapes of isobars near their point of inflexion between quasi-stationary troughs and ridges in a general westerly flow.

Equation (20) does not permit of the shapes of stream-lines near their points of inflexion indicated below in diagrams ( $\gamma$ ) and ( $\delta$ ), for an easterly flow :



Stream-lines of the shapes ( $\gamma$ ) and ( $\delta$ ) are therefore inconsistent with the conditions specified at the beginning of this example. One or more of those conditions must be relaxed.

But although the shapes ( $\alpha$ ) and ( $\beta$ ) are familiar in the case of isobars, the order of magnitude of  $\frac{\partial \kappa}{\partial s}$  where  $\kappa=0$  appears to be often of the order of  $\frac{-1}{50,000} \text{ miles}^{-2}$ , or 10 times as big as the value of  $\frac{\partial \kappa}{\partial s}$  required by equation (23).

In such cases, one or more of the conditions specified in this example must be relaxed.

In the next example, the condition of horizontal flow throughout a neighbourhood of the place is relaxed.

(iv) Unsaturated air, under dry-adiabatic conditions; gradient flow throughout a horizontal neighbourhood of the place, geostrophic at the place itself.

By hypothesis  $\kappa=0$ .

Taking the same values as in example (i), and in addition

$$\frac{\partial p}{\partial t}=0, \quad \frac{\partial \kappa}{\partial s}=-\frac{1}{50,000},$$

equation (20) gives

$$\begin{aligned} \frac{\partial w}{\partial z} &= -0.08 \text{ m.p.h. per vertical mile} \\ &= -2 \text{ cm. per sec. per vertical km.} \end{aligned} \quad (26)$$

Taking instead of  $\frac{\partial p}{\partial t}=0$ ,  $\frac{\partial p}{\partial t}=\pm 2$  mb. per hour, affects the value of  $\frac{\partial w}{\partial z}$  given by equation (26) by only 2 per cent.

#### 10. Further Consideration of Equation (18).

Equation (18), obtained in section 6, is repeated below.

Taking magnitudes typical of those normally experienced :

$$p=700 \text{ mb.}, \quad \frac{\partial p}{\partial t}=1 \text{ mb. per hour}, \quad G=40 \text{ m.p.h.}, \quad \kappa=-\frac{1}{300} \text{ miles}^{-1},$$

$$\frac{\partial G}{\partial s}=5 \text{ m.p.h. per 100 miles}, \quad \frac{\partial \kappa}{\partial s}=-\frac{1}{200,000} \text{ miles}^{-2}, \quad \phi=45^\circ,$$

$$G_{S \rightarrow N}=30 \text{ m.p.h.}, \quad \frac{\partial x}{\partial s}=-\frac{1}{40,000} \text{ miles}^{-1}, \quad T=270^\circ \text{ A.},$$

$$\frac{\partial T}{\partial s}=-1^\circ \text{ A. per 100 miles.}$$

The order of magnitude of each term in equation (18) is indicated immediately below the term to which it applies (the actual values have been multiplied by 10,000) :

$$\begin{aligned} \frac{1}{\gamma p} \frac{\partial p}{\partial t} &= \frac{1}{G\kappa+\lambda} \kappa G \frac{\partial G}{\partial s} + \frac{1}{G\kappa+\lambda} G^2 \frac{\partial \kappa}{\partial s} + \frac{1}{G\kappa+\lambda} \frac{2w \cos \phi}{R_E} G_{S \rightarrow N} \\ &\quad - \frac{3}{5} G \frac{\partial x}{\partial s} - \frac{G}{T} \frac{\partial T}{\partial s} - \frac{\partial w}{\partial z}. \end{aligned} \quad (18)$$

$$10 = -280 - 340 + 120 + 6 + 15 + 490.$$

Equation (18) thus reduces effectively to

$$\frac{\partial w}{\partial z} = \frac{1}{G\kappa+\lambda} \left\{ \kappa G \frac{\partial G}{\partial s} + G^2 \frac{\partial \kappa}{\partial s} + \frac{2w \cos \phi}{R_E} G_{S \rightarrow N} \right\}. \quad (27)$$

Wherever the conditions specified in section 2 are fulfilled, equation (27) can be used to obtain the value of  $\frac{\partial w}{\partial z}$ .

Moreover, equation (27) is still true to the same degree of approximation if the dry-adiabatic condition is relaxed, for if the air is unsaturated

$$p = R \left( 1 + \frac{3}{5} x \right) \rho T.$$

Therefore 
$$\frac{1}{\rho} \frac{\partial \rho}{\partial t} = \frac{1}{p} \frac{\partial p}{\partial t} - \frac{1}{T} \frac{\partial T}{\partial t} - \frac{3}{5} \frac{\partial x}{\partial t}, \quad \dots \quad (28)$$

with sufficient approximation.

But if the wind is gradient throughout a horizontal neighbourhood,

$$\begin{aligned} \frac{1}{\rho} \frac{\partial \rho}{\partial t} &= - \frac{\nabla(\rho \mathbf{V})}{\rho} \\ &= - \frac{\nabla(\rho \mathbf{G})}{\rho} - \frac{\partial w}{\partial z}. \quad \dots \quad (29) \end{aligned}$$

Thus (28) and (29) give

$$\frac{1}{p} \frac{\partial p}{\partial t} - \frac{1}{T} \frac{\partial T}{\partial t} - \frac{3}{5} \frac{\partial x}{\partial t} = - \frac{\nabla(\rho \mathbf{G})}{\rho} - \frac{\partial w}{\partial z}, \quad \dots \quad (30)$$

and consideration of the order of magnitude of the terms on the left-hand side shows that equation (30) reduces effectively to equation (27).

It is intended to investigate statistically the applicability of equation (27).

The author is indebted to Mr. C. H. B. Priestley for much helpful criticism, and to the Director of the Meteorological Office for permission to publish this paper,

#### SUMMARY.

1. Except in high latitudes, horizontal geostrophic flow with an appreciable northerly or southerly component implies a pressure-tendency much larger than normally observed. This excessive pressure-tendency can be moderated or nullified by assuming throughout a horizontal neighbourhood gradient flow geostrophic only at the place, with slight curvature of the air-trajectory up and down wind, and/or a slight gradient of vertical velocity.

2. Gradient flow usually implies terms in the formula for the divergence much larger than the term involving the pressure-tendency.

3. Neglecting small terms, including the pressure-tendency, gradient flow throughout a horizontal neighbourhood, of unsaturated air, implies a gradient of vertical velocity, for which a formula is obtained.

#### BIBLIOGRAPHY.

- (1) H. Jeffreys, "On Travelling Atmospheric Disturbances," *Phil. Mag.* xxxvii. (1919).
- (2) S. Petterssen, 'Weather Analysis and Forecasting,' pp. 228-231.
- (3) D. Brunt, 'Physical and Dynamical Meteorology,' pp. 310-312, etc.
- (4) C. J. Boyden, *Q. J. Roy. Met. Soc.* pp. 85-91.
- (5) Hewson and Longley, 'Meteorology, Theoretical and Applied,' pp. 127, 128, 289-293 and pp. 356-358.
- (6) R. C. Sutcliffe, *Q. J. Roy. Met. Soc.* lxiv. p. 496 (1938).
- (7) H. Jeffreys, *Q. J. Roy. Met. Soc.* xlviii. 1922, pp. 29-46.
- (8) H. Lamb, 'Hydrodynamics,' (4th edn.), pp. 318-320,



LXXIV. *A Criticism of Two Well-known Theorems on Most Efficient Propellers.*

By V. D. NAYLOR, M.Sc.\*

[Received January 8, 1946.]

A MOST efficient propeller is defined as a propeller which, at a given forward speed  $V$  and a given angular speed  $\Omega$ , requires the least amount of input power to produce a given output power. The working fluid will be assumed non-viscous. There are two theorems in connection with a best propeller (B.P.), which we refer to as Theorems A and B.

THEOREM A.

The inflow factors  $a$  and  $a'$  for a propeller working most efficiently are connected by the relation

$$\frac{1+a}{1-2a'} + \frac{a'x^2}{1+2a} = \text{constant},$$

where  $x = \frac{r\Omega}{V}$

THEOREM B.

If a best propeller is lightly loaded, in which case  $\eta_e \doteq 1$  and  $a$  and  $a'$  are both small, then

$$\frac{1-a'}{1+a}$$

is approximately constant along the blade.

Glauert † gives a proof of Theorem A, which he calls an exact theorem by using the simple momentum theory, and so,  $a$  and  $a'$  denote *average* inflow factors—mean values for the annular disc element  $2\pi r\delta r$ —to which they refer. He deduces Theorem B from A with the help of the extra assumption that the propeller is lightly loaded, so that Theorem B is the approximate statement of Theorem A when the propeller is lightly loaded.

Glauert's proof ignores the periodicity of flow. The object of this paper is to show what takes the place of A and B when periodicity of flow is admitted. Consider a propeller advancing into still air with speed  $V$ , viscous drag being neglected. The only wasted power is the axial K.E. and the rotational K.E. given to the slipstream. The trailing vortices will produce induced velocities in the fluid in the immediate neighbourhood of the blade element, which we denote by  $a_0V$  and  $a_0'r\Omega$

\* Communicated by the Author.

† 'Aeroplane Propellers,' by H. Glauert, p. 197.

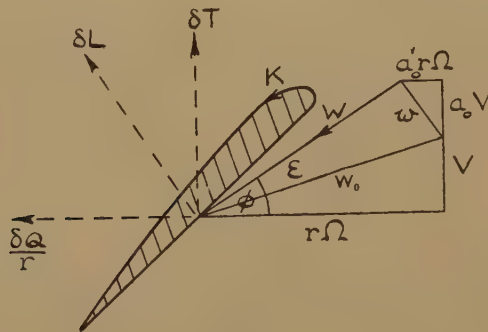
(fig. 1). Generally, the inflow factors at a point will be functions of the polar co-ordinates of that point relative to an initial line lying along the blade, so that we may write

$$a=f_1(r, \psi); \quad a'=f_2(r, \psi),$$

and  $f_1(r, 0)$ ,  $f_2(r, 0)$  are what we have denoted by  $a_0$  and  $a_0'$ . If we suppose that the inflow factors are constant round the element  $2\pi r\delta r$ , as we do when we assume non-periodicity of flow, we shall denote their values by  $a_0$  and  $a_0'$ .

The velocity of the air relative to the blade,  $W$ , is got by compounding  $a_0'r\Omega$  and  $a_0V$  with  $r\Omega$  and  $V$ , the velocity components of the blade relative to earth. The lift  $\delta L$  on  $c\delta r$  will be perpendicular to  $W$  and the corresponding lift coefficient  $C_L$  will be that for two-dimensional flow. The component of  $\delta L$  along the direction of  $W_0$  is what is known as the induced drag.

Fig. 1.



Let  $K$  be the circulation round the element  $c\delta r$ , and let  $B$  be the number of blades.

We have the following well-known results:—

$$\delta L = \rho K W \delta r = C_L \rho / 2 c \delta r W^2; \quad \therefore K = \frac{c C_L W}{2} \quad . \quad . \quad . \quad (1)$$

$$\delta T = B \rho K r \Omega (1 - a_0') \delta r; \quad T = \int_0^R B \rho K r \Omega (1 - a_0') dr \quad . \quad . \quad . \quad (2)$$

$$\delta Q = B \rho K V (1 + a_0) r \delta r; \quad Q = \int_0^R B \rho K V r (1 + a_0) dr \quad . \quad . \quad . \quad (3)$$

$$\eta_r = \frac{\delta T \cdot V}{\delta Q \cdot \Omega} = \frac{B \rho K V \Omega r (1 - a_0') \delta r}{B \rho K V \Omega r (1 + a_0) \delta r} = \frac{1 - a_0'}{1 + a_0} \quad . \quad . \quad . \quad (4)$$

Input—output

$$\left. \begin{aligned} &= \Omega \delta Q - V \delta T, \\ &= B \rho K V \Omega r (1 + a_0) \delta r - B \rho K V \Omega r (1 - a_0') \delta r, \\ &= B \rho K V \Omega r (a_0 + a_0') \delta r \end{aligned} \right\} \quad . \quad . \quad . \quad (5)$$

Power given to slipstream by  $\delta T$  and  $\delta Q$

$$\left. \begin{aligned} &= \delta T \cdot a_0 V + \delta Q \cdot a_0' \Omega, \\ &= B \rho K r \Omega (1 - a_0') a_0 V \delta r + B \rho K V (1 + a_0) r \delta r a_0' \Omega, \\ &= B \rho K V \Omega r (a_0 + a_0') \delta r. \end{aligned} \right\} \quad (6)$$

The above two are equal, showing that we have an exact balance of power, as we expect, expressed by

$$\Omega \delta Q - V \delta T = \delta T \cdot a_0 V + \delta Q \cdot a_0' \Omega.$$

#### THE PROBLEM.

We have

$$\left. \begin{aligned} TV &= B \rho V \Omega \int_0^R r K (1 - a_0') dr \\ Q \Omega &= B \rho V \Omega \int_0^R r K (1 + a_0) dr \end{aligned} \right\}, \quad (7)$$

and

we use  $\Delta a_0$ ,  $\Delta a_0'$  to denote changes in the form of the functions  $a_0$  and  $a_0'$  and  $\Delta K$ ,  $\Delta(TV)$  to denote the consequent changes in  $K$  and  $(TV)$  as is understood in the calculus of variations.

We have

$$\left. \begin{aligned} \Delta(TV) &= B \rho V \Omega \int_0^R r \{ \Delta K (1 - a_0') - K \Delta a_0' \} dr, \\ \Delta(Q \Omega) &= B \rho V \Omega \int_0^R r \{ \Delta K (1 + a_0) + K \Delta a_0 \} dr \end{aligned} \right\} \quad (8)$$

By a known theorem in the calculus of variations,  $TV$  will be a maximum for a given value of  $Q \Omega$  if the ratio of the integrands in  $\Delta(TV)$  and  $\Delta(Q \Omega)$  is independent of  $r$ , i. e. if

$$\frac{(1 - a_0') \Delta K - K \Delta a_0'}{(1 + a_0) \Delta K + K \Delta a_0} \quad (9)$$

is constant along the blade, then we have a B.P.

We shall call this the *general condition for a best propeller*.

Suppose now that

$$\frac{1 - a_0'}{1 + a_0}$$

is constant along the blade and equal to, say,  $A$ , we have  $1 - a_0' = A(1 + a_0)$  ;  
 $-\Delta a_0' = A \Delta a_0$ . Then

$$\frac{(1 - a_0') \Delta K - K \Delta a_0'}{(1 + a_0) \Delta K + K \Delta a_0} = \frac{A(1 + a_0) \Delta K + K A \Delta a_0}{(1 + a_0) \Delta K + K \Delta a_0} = A,$$

a constant along the blade, and we have therefore a B.P.

If

$$\frac{1 - a_0'}{1 + a_0}$$

is not constant along the blade, but equal to, say,  $\phi(r)$ , then

$$1 - a_0' = \phi(r)(1 + a_0); \quad -\Delta a_0' = \phi(r) \cdot \Delta a_0 \text{ and}$$

$$\frac{(1 - a_0')\Delta K - K\Delta a_0'}{(1 + a_0)\Delta K + K\Delta a_0}$$

reduces to  $\phi(r)$ , which is not constant along the blade, and so we have not a B.P.

Hence we have the theorem :—

#### THEOREM A'.

*A propeller is a B.P. if, and only if,*

$$\frac{1 - a_0'}{1 + a_0}$$

*is constant along the blade.*

From (4) it follows that a propeller is most efficient if, and only if, the efficiency of every element along the blade is the same.

In these circumstances, the efficiency of the blade ( $\eta$ ) is equal to the constant efficiency of the elements, and we may say that, for a B.P.,

$$\eta = \frac{1 - a_0'}{1 + a_0} \quad \dots \dots \dots (10)$$

When periodicity of flow is admitted, Theorem A' is the exact condition for a best propeller; there is no question of the propeller being lightly loaded, as this assumption has not been made anywhere. Theorem A' should be contrasted with Theorem B, which gives the same condition as an approximate result which is only proved to hold when the flow is assumed to be non-periodic and the propeller is lightly loaded. The seeming paradox disappears when we recollect that the  $a_0$  and  $a_0'$  of Theorem A' are inflow factors at a point and the  $a$  and  $a'$  of Theorem B are mean values taken over the annulus  $2\pi r \delta r$ .

To sum up: with considerably fewer assumptions we have deduced a theorem of wider generality.

Let us now inquire how Theorem B enters into the scheme of things. We shall need one or two preliminary theorems. •

*The Velocity in the Wake = 2 × the Velocity at the Disc.*

We can easily deduce this result if we assume non-periodicity.

Consider a cylindrical element of the slipstream  $2\pi r \delta r$  and let the velocity at the disc be  $a_c V$  at all points of  $2\pi r \delta r$ . Let the velocity at all points in the annulus at the wake be  $b_c V$  (fig. 2).

Then :—

Mass of air attacked per sec. by  $2\pi r \delta r = 2\pi r \delta r \rho (V + a_c V)$ .

Gain of momentum per sec. =  $2\pi r \delta r \rho (V + a_c V) b_c V$ .

If  $\delta T$  denotes the thrust contribution made by  $2\pi r \delta r$ , then

$$\delta T = 2\pi r \delta r \rho (V + a_c V) b_c V.$$

Work done per sec. by  $\delta T$  on the air =  $\delta T \cdot a_c V$

$$= 2\pi r \delta r \rho (V + a_c V) \cdot b_c V \cdot a_c V.$$



Linear K.E. given per sec. to S.S. =  $\frac{1}{2}(\text{mass per sec.})(\text{Vel.})^2$

$$= \frac{1}{2} \cdot 2\pi r \delta r \rho (V + a_c V)(b_c V)^2.$$

$$\therefore 2\pi r \delta r \rho (V + a_c V) \cdot b_c V \cdot a_c V = 1/2 \cdot 2\pi r \delta r \rho (V + a_c V) (b_c V)^2.$$

$$\therefore b_c = 2a_c.$$

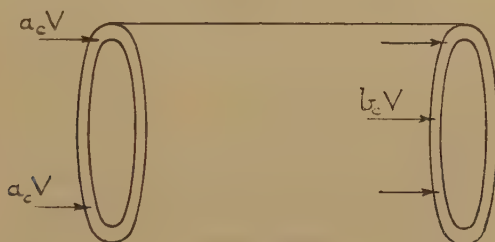
Hence, with the assumption made we may express the axial velocity of the air at the disc and in the wake by

$$a_c V \text{ and } 2a_c V. \quad (11)$$

Similarly, by assuming that the angular momentum of the slipstream is caused by the torque which the element  $2\pi r\delta r$  of the disc exerts on the air and that the K.E. of rotation is caused the same way, we may show that the angular velocity of the air at the disc and in the wake may be expressed by

$$a_c' \Omega \text{ and } 2a_c' \Omega. \quad (12)$$

Fig. 2.



The assumption on which the proof rests is that the flow is non-periodic and this would be the case if the propeller had an infinite number of blades. With this assumption the expressions for thrust and torque are

$$\left. \begin{aligned} \delta T &= 4\pi r \delta r \rho V^2 (a_c + a_c'^2) \\ \delta Q &= 4\pi r^3 \rho V \Omega \delta r (1 + a_c) a_c' \end{aligned} \right\} \dots \dots \dots (13)$$

It is usual to assume that in ALL cases, the velocity in the wake is twice the velocity on the disc, but we shall not delay to examine how this assumption affects the rationalization of propeller theory.

### Relation between the Interference Factors.

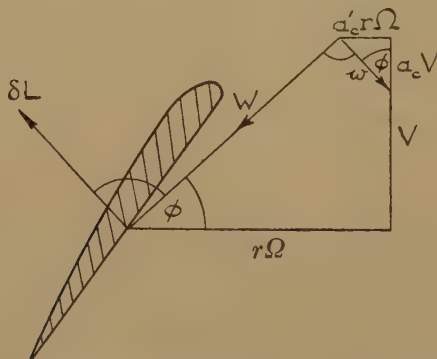
Denote the velocity at the disc which has components  $a_c V$ ,  $a_c' r \Omega$  by  $\vec{w}$  (fig. 3). The velocity in the wake, with the assumptions made above, would be  $2\vec{w}$ . The change in the velocity  $2\vec{w}$  is produced by the action of  $\delta L$  on the air and so  $2\vec{w}$  is parallel to  $\delta L$ , and therefore perpendicular to  $W$ . Hence  $\vec{w}$  is  $\perp \vec{W}$ , and the angles marked  $\phi$  are equal.

As a result we have

$$\begin{aligned} \tan \phi &= \frac{V+a_cV}{r\Omega-a'_c r\Omega} = \frac{a_c r\Omega}{a_c V}, \\ \therefore a_c+a_c^2 &= (a'_c-a'_c)^2 X^2, \\ \text{where} \qquad \qquad X &= \frac{r\Omega}{V}. \qquad \qquad \qquad (14) \end{aligned}$$

Whenever we use this result we are assuming that the inflow factors are constant round  $2\pi r\delta r$ , or that there is an infinite number of blades (whether the propeller is lightly loaded or not).

Fig. 3.



It is to be noted (fig. 1) that if  $a_0$  and  $a'_0$  are small (light loading), then  $\epsilon$  is small and  $w$  is approximately perpendicular to  $W$ .

$$\therefore \tan \phi = \frac{V(1+a_0)}{r\Omega(1-a'_0)} \text{ and } \tan \phi = \frac{a'_0 r\Omega}{a_0 V},$$

giving

$$a_0+a_0^2 = (a'_0-a'_0)^2 X^2. \qquad \qquad \qquad (15)$$

We may use this expression when  $a_0$  and  $a'_0$  are small, *i.e.* when there is light loading (whether there is an infinite number of blades or not).

*Non-periodic Flow.*

Equating  $\delta T$ , as found from the momentum and vortex theories (13) and (2), we get

$$\begin{aligned} 4\pi r \rho \delta r V^2 (1+a_c) a_c &= B K \rho r \Omega (1-a'_0) \delta r, \\ \therefore K &= \frac{4\pi V^2}{B \Omega} \cdot \frac{a_c+a_c^2}{1-a'_0}. \qquad \qquad \qquad (16) \end{aligned}$$

This is a hybrid formula involving both periodic and non-periodic flow. If we replace  $o$  by  $c$  we get a result which holds when non-periodic flow is assumed, *i.e.* when there is an infinity of blades. We get

$$K = \frac{4\pi V^2}{B \Omega} \cdot \frac{a_c+a_c^2}{1-a'_c}. \qquad \qquad \qquad (17)$$



and if we assume non-periodic flow we may write

$$\eta_r = \frac{1 - a_c'}{1 + a_c'}$$

Using (14), we get

$$\eta_r = \frac{a_c}{a_c' X^2}$$

$$\therefore a_c = X^2 \eta_r a_c' = X^2 \eta_r [1 - \eta_r (1 + a_c)]$$

$$\therefore a_c = \frac{X^2 \eta_r (1 - \eta_r)}{1 + X^2 \eta_r^2} \quad \dots \dots \dots (20)$$

and

$$a_c' = \frac{a_c}{X^2 \eta_r} = \frac{1 - \eta_r}{1 + X^2 \eta_r} \quad \dots \dots \dots (21)$$

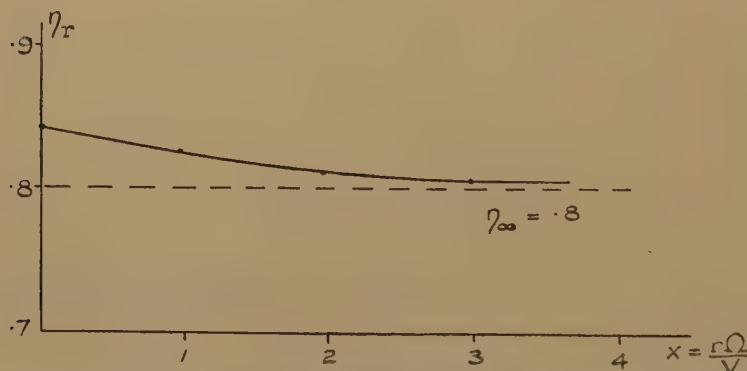
Substituting for  $a_c$  and  $a_c'$  in (B') we get

$$\frac{X^2 (1 - \eta_r)}{1 + X^2 \eta_r (2 - \eta_r)} + \frac{1 + X^2 \eta_r}{2 \eta_r - 1 + X^2 \eta_r^2} = \text{constant} = \frac{3 - 2 \eta_\infty}{\eta_\infty (2 - \eta_\infty)} \quad (\text{N.P.*}),$$

where  $\eta_\infty$  denotes the efficiency of an element at infinity, i. e.  $\eta_r \rightarrow \eta_\infty$  when  $X \rightarrow \infty$ .

A graph is given (fig. 4) showing  $\eta_r$  against  $X$  when  $\eta_\infty = 0.8$ . This graph shows the deviation from constancy in the efficiency of an element

Fig. 4.



Graph showing efficiency of element of a best propeller with and without periodicity.

of a best propeller consequent on the assumption of non-periodicity. It is to be stressed that nowhere have we made the assumption of light loading, or that  $a$  and  $a'$  are small and  $\eta_r \doteq 1$ .

Theorem B' is to be contrasted with Theorem A; we have deduced B' as an approximate theorem for the actual flow: Glauert quotes it as an exact theorem.

Let us sum up the position.

\* N.P. is inserted to indicate that the result depends on the supposition of non-periodicity, usually the presence of the suffix  $c$  is sufficient to indicate this.



We have shown that

$$\frac{1-a_0'}{1+a_0} = \text{constant}$$

is an exact condition for a B.P.

$$\frac{1+a_c}{1-2a_c'} + \frac{a_c'}{1+2a_c} X^2 = \text{constant}$$

is an approximate condition assuming that periodicity is ignored.

In this we have not made any assumption of light loading.

Glauert gives the results

$$\frac{1+a_c}{1-2a_c'} + \frac{a_c'}{1+2a_c} X^2 = \text{constant}$$

is an exact condition for a B.P.

$$\frac{1-a_c'}{1+a_c} = \text{constant}$$

is an approximate condition for a B.P. when the propeller is lightly loaded.

It is difficult to see how the assumption of non-periodicity, which is made in the axial momentum theory, could lead to a result which could be called exact as is done by Glauert. It is easy to see how Glauert's second theorem follows from the first with the extra assumption of light loading, thus :

$$\frac{1+a_c}{1-2a_c'} + \frac{a_c'}{1+2a_c} X^2 = \text{constant},$$

gives 
$$\frac{1-a_c}{1-2a_c'} + \frac{a_c'}{1+2a_c} \cdot \frac{a_c(1+a_c)}{a_c'(1-a_c')} = \text{constant (using 14),}$$

$$\frac{1+a_c}{1-a_c'} \left( \frac{1+a_c'}{1+2a_c'} + \frac{a_c}{1-2a_c} \right) = \text{constant},$$

$$\frac{1+a_c}{1-a_c'} \doteq \text{constant}$$

if  $a_c$  and  $a_c'$  are small.

But we do not agree that the first theorem is exact.

This really completes the limited object of this paper, which was to inquire into the significance of Theorems A and B in a rational development of propeller theory. It may be worth while commenting on Glauert's later deduction of best-propeller theorems from the vortex theory.

He writes (p. 254) :  $\Delta T = \rho \Delta K (\Omega - \omega) r dr.$

$$\Delta Q = \rho \Delta K (V + v) r dr.$$

$$\frac{\rho \Delta K (\Omega - \omega) r}{\rho \Delta K (V + v) r}$$

is constant along the blade, where  $r\omega$  and  $v$  are the component velocities in the wake which we should denote by  $2a'r\Omega$  and  $2aV$ .



Suppose the helix to which OC is tangent is moved axially backward with velocity  $\overline{BE}$  and denote BE by  $v_1$ .

We have

$$BE = AE - AB = OA \tan \phi - AB = r\Omega \cdot \frac{V}{r\Omega} \left( \frac{1+a_0}{1-a_0'} \right) - V,$$

$$\therefore v_1 = BE = V \left( \frac{1+a_0}{1-a_0'} - 1 \right).$$

For a B.P.,  $\frac{1+a_0}{1-a_0'}$  is constant along the blade, and therefore  $v_1$  is constant along the blade and the whole helicoid will move backwards with axial velocity  $v_1$ . It remains to see in what circumstances the axial motion of this helicoid will set up in the fluid what we know to be the interference velocity due to the motion of the propeller, namely  $\overline{CB}$ .

In what circumstances is  $\overline{CB}$  equal to  $\overline{EB} \sin CEB$ , for the velocity of the fluid due to the motion of the helicoid is equal to the velocity of the helicoid normal to itself.

This requires that CB should be perpendicular to OC, which we have seen will be the case if we assume non-periodicity, *i. e.* if we assume  $a$  and  $a'$  are constant over the annulus  $2\pi r \delta r$ , which is equivalent to assuming an infinity of blades. In these circumstances we may replace  $a_0$  and  $a_0'$  by  $a_c$  and  $a_c'$ , and we may say that the interference velocities due to the motion of a best propeller with an infinite number of blades can be exactly produced by the axial motion of a rigid helicoid with axial velocity

$$V \left( \frac{1+a_c}{1-a_c'} - 1 \right),$$

this being the same for all points of the blade. In actual practice, this is not an exact theorem because no propeller has an infinity of blades. If we try to go further and deduce theorems about the periodic flow of air in the slipstream, we are up against the logical difficulty of deducing results for periodic flow from results founded on the supposition of non-periodic flow.

The question arises, can we approximately reproduce the interference velocities by the axial motion of such a helicoid? The answer is yes, provided  $a$  and  $a'$  are small, *i. e.* provided the propeller is lightly loaded, for when this is so CB is sensibly perpendicular to OC.

We sum up the results thus:—

The axial motion of the helicoid with velocity

$$v_1 = V \left( \frac{1+a}{1-a'} - 1 \right)$$

will reproduce the induced velocity due to the motion of the propeller (1) accurately when the  $a$ 's are assumed constant round  $2\pi r \delta r$ , *i. e.* when there is an infinity of blades, or (2) approximately when the  $a$ 's are small, *i. e.* when there is light loading. This  $v_1$  will do duty for all elements

when the propeller is a B.P. In either case we have only an approximation to the actual flow.

It appears from the above that as we decrease the number of blades we must invoke more and more the supposition of light loading to justify us in saying that the interference velocities can be reproduced by the axial motion of a rigid helicoid. With a large number of blades it does not matter whether the propeller is lightly loaded or not. But with a small number of blades the propeller must be lightly loaded.

This is about as far as we can go with the rational development along simple lines. If we try and extend the idea of helicoidal flow to the motion in the wake we meet difficulties, one of which is that we have to make use of the assumption that the velocity in the wake is double the velocity on the disc which, so far as this paper goes, is only proved when there is an infinity of blades. On top of all this, a further question arises: to what extent is the hypothesis of light loading an overriding assumption doing away with the need for assuming a best propeller? Referring to fig. 1: if  $a_0$  and  $a_0'$  are small, then  $\epsilon$  is small and  $W$  is therefore tangential to a surface which only makes a small angle with the truly helicoidal surface to which  $W_0$  is tangential. So that we may say that with light loading the air at the disc is *approximately* incident along a helicoidal surface, and in this statement there is no assumption made of a best propeller.

This is mentioned because the author has shown elsewhere\* that experimental evidence exists for believing that, for a certain family of propellers, the air is incident along helicoidal sheets from the value of  $J$  corresponding to zero-thrust to the value just below maximum efficiency, and these propellers are not best propellers as defined at the opening of this paper.

This indicates that the assumption of a best propeller on top of the assumption of light loading is superfluous in this case.

Ministry of Aircraft Production,  
Millbank, London.

## LXXV. *Determination of Refractive Index of a Liquid by Newton's Rings.*

By R. L. WOOLLEY, B.A., Trinity College, and  
G. F. C. SEARLE, Sc.D., F.R.S., Cavendish Laboratory, Cambridge †.

[Received August 2, 1945.]

### § 1. INTRODUCTION.

IN his experiments on the rings formed between a lens and a plate of glass, Newton found that, with light contact between lens and plate, the diameters of the rings were such that the distance  $t_n$  between lens and

\* "The Torque Coefficient for a Family of Propellers," 'Aircraft Engineering,' vol. xvii. No. 198 (Aug. 1945).

† Communicated by the Authors.



plate, corresponding to the  $n$ th ring, counting outwards from the centre, was less, in the ratio of about three to four, when there was water between lens and plate than when there was air between them. "Perhaps," he says, "it may be a general rule" that, in modern words, the distance  $t_n$  is inversely proportional to the refractive index of the medium between the lens and the plate. He found that the coloured rings were fainter with water than with air between the lens and plate ('Optics,' Book II. Part I, Obs. 10).

In the ordinary method, light is reflected vertically down upon the system of lens and horizontal plate by a transparent reflecting plate, set at  $45^\circ$ , and the rings are viewed through this reflector by a vertical microscope. With sodium light, and with water between lens and plate, the contrast between the "bright" and "dark" rings is so small that no accurate measurements of their diameters are possible. In this form, the experiment is disappointing.

The difficulty is due to unwanted reflections. We suppose that the horizontal plate of glass rests on the upper, convex, surface of a lens. Light, sent down by the oblique reflector, is copiously reflected from A, the upper surface of the plate, and from D, the lower surface of the lens, because the changes of refractive index at A and D are large. At the interface B between plate and water and at the interface C between water and lens, the change of index is small and little light is reflected. The light available for the formation of rings by reflexions at B and C is therefore small, and, in consequence, the rings are practically swamped by the light reflected at A and D. Some improvement is obtained by coating the face D of the lens with black varnish, which reduces the reflexion at D. For students who wish to determine the wave-length, it is a disadvantage that the radius of curvature of the face C of the lens cannot now be found by Boys's method. The radius can, however, be measured, with a little more trouble, by the method of Experiment 20 in the 'Experimental Optics' of G. F. C. Searle. A much greater improvement results if, as in § 2, no light reflected at A ever enters the microscope.

## § 2. METHOD.

A cubical glass block J (figs. 1, 2), such as is used in elementary optical experiments\*, rests on a base UU, shown broken in fig. 1. The base has two feet and a levelling screw I, and these stand on the horizontal slab or table V, which supports the horizontal track of the vertical travelling microscope M. Two faces of the block are parallel to the plane of fig. 1. The track of M is perpendicular to that plane.

Let a vertical ray, corresponding to the axis of M, meet the face A of the block J at G and be refracted along GH to meet the face B of J at H.

---

\* The faces of one block were nearly plane and gave large rings when tested against a flat plate. But the rings were jagged, somewhat like the teeth of a circular saw. The faces, after their grinding, were probably "polished" on a buffing wheel. Mr. A. H. L. Dunn, who made the apparatus, re-polished the faces A, B with rouge on a wax polisher, and obtained satisfactory surfaces.

Let the ray reflected at H meet A in K. By the screw I, the tilt of J is adjusted so that GK is not more than 5 or 6 mm. A right-angled prism P sits on the face A; its hypotenuse face makes  $45^\circ$  with A. The rectangular face, which is now perpendicular to A, is approximately coplanar with the face X of the block. The base UU is then adjusted on the table V relative to the microscope, so that the shortest distance from G to the line, in which the (silvered) hypotenuse face of P meets A, is not more than 2 or 3 mm. If light from a suitably placed source Q fall on the "vertical" face of P, a ray KH will be reflected along HG and will enter the microscope along its axis. If the thickness AB of the block be 5 cm., and if GK be less than 0.6 cm., the angle of incidence at H is less than  $3^\circ 30'$ .

If the convex face C of a lens L touch the face B of the block at H, a system of rings, with centre at H, will be seen when the microscope is suitably focused. They are practically "located" on the face B of the block. They may, of course, be seen without the aid of the microscope.

Fig. 1.

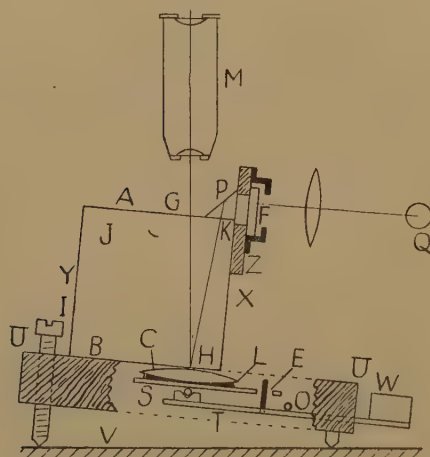
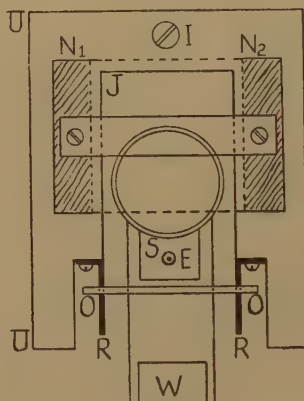


Fig. 2.



It is essential that the lens should touch the block at H. To secure this by any screw arrangement would be difficult, because the radius of curvature of C is of the order of 100 cm. A simple device meets the case.

The lower face of the lens is coated with black varnish. The lens is fixed with wax to a plate S. A cycle ball,  $\frac{1}{8}$  in. in diameter, is soldered to S as near as may be to where the axis of L meets S. The ball is seated in a conical hole in the top of a short pillar rising from the lever T. The lever is supported by an axle OO (seen end-on at O in fig. 1), which rests on two rails R, R (fig. 2) attached to the base UU; they are parallel to the plane of fig. 1 and to the face B of the block. A weight W makes L press against B. The lens, on its ball seating, finds its own point of contact. By adjusting the axle on the rails, we can bring the centre of the rings to any desired point in the field of view of M. The lens is large enough to

project a little beyond the face X of the block. A vertical wire projecting from T and passing, with ample clearance, through a hole E in one end of S, keeps S under control during the assembly. The E-end of S serves as a tag by which L can be mounted on an optical bench for the determination of the radius of curvature of the face C.

The base UU is shaped as in fig. 2, with a gap to accommodate the lens end of the lever system under the block J. The "plan" of J is shown by the dotted line. Further details are shown in fig. 2; the prism and its fittings are omitted. The glass block J is closely flanked by two wooden blocks  $N_1$ ,  $N_2$  fixed to the base UU, and is clamped to UU by a bridge piece screwed to  $N_1$  and  $N_2$ .

The prism P and the face X of J butt against a plate Z (fig. 1), which is screwed to  $N_1$ ,  $N_2$ . A hole in Z allows light from a source Q to reach P. A filter F, for isolating the green mercury line ( $\lambda = 5.461 \times 10^{-5}$  cm.), can be placed in a fitting carried by Z. A suitable strap screwed to Z keeps P in place.

In place of the two blocks  $N_1$ ,  $N_2$  of fig. 2, a single block N, fixed to UU and flanking the face Y of J, may be used. Bolts passing through N and through Z serve to clamp J against N. The gap in UU can now be wider and the lens can be of larger area—an advantage if the radius of curvature of the face C is to be found.

The glass block merely acts as a window through which the rings are viewed. The rays are so nearly normal to the faces A, B that the astigmatic effects due to obliquity are negligible.

A sodium flame is placed so that the rings are seen. A suitable lens is then interposed between the flame and P to secure the best illumination. If a mercury lamp be used, the filter F is placed in its seating on Z. With air between B and C, the ring system is brilliant, particularly with the green mercury light. If a drop of water be run on to the face X from a pipette, some, guided by the part of the lens projecting beyond X, will creep in between B and C. The scale of the ring system will diminish; the pattern, though not so brilliant as before, will be bright enough and crisp enough for good measurement.

The rays which have suffered two reflections at C are of small intensity, but are strong enough to make the dark rings narrow relative to the bright intervals between them.

If the faces B and C of block and lens be not clean where they touch, the central spot may change from bright to dark or *vice versa* when the slab V is tapped.

The diameters of the rings for air and for water are measured. Let them be  $a_1, a_2 \dots$  for air, counting outwards from the centre, and  $w_1, w_2 \dots$  for water. If  $\mu$  be the refractive index of water,  $\mu = (a_n^2 - a_1^2) / (w_n^2 - w_1^2)$ . The readings may be arranged as in § 3.

The incidence at H is so nearly normal that the conventional theory of Newton's rings may be applied without question.

The apparatus may be used for a modification of an experiment of Newton ('Optics,' Book II, Part I, Obs, 13). He observed the rings



formed when the light falling on the plate and lens came from a definite (small) part A of a sheet of white paper. A spectrum was formed on the paper by a prism. When the prism was turned so that the colour falling on A was changed from blue to red, the rings expanded. "I found the Circles which the red Light made to be manifestly bigger than those which were made by the blue and violet. And it was very pleasant to see them gradually swell or contract accordingly as the Colour of the Light was changed." In our method, a lamp with a straight filament, a slit, a lens and a prism are used to form a spectrum, as pure as possible, across the opening in the plate Z. The prism is on a turning table about 100 cm. from Z. When the table is turned, the spectrum moves across Z and the wave-length of the light forming the rings is changed. The change of scale noted by Newton is well seen.

### § 3. PRACTICAL EXAMPLE.

Observations by R. L. W. are intended to show the sort of result a student may obtain in half an hour's work. The diameter of the  $n$ th ring, counting outwards from the centre, is  $a_n$  for air and  $w_n$  for water between lens and block.

$n$	$a_n$	$a_n^2$	$a_n^2$ calc.	$w_n$	$w_n^2$	$w_n^2$ calc.
	mm.	mm. <sup>2</sup>	mm. <sup>2</sup>	mm.	mm. <sup>2</sup>	mm. <sup>2</sup>
12	4.795	22.992	23.089	4.165	17.347	17.2574
10	4.385	19.228	19.196	3.775	14.251	14.3068
8	3.920	15.366	15.302	3.365	11.323	11.3562
30		57.586			42.921	
6	3.385	11.458	11.409	2.905	8.439	8.4056
4	2.735	7.480	7.516	2.320	5.382	5.4550
2	1.900	3.610	3.623	1.595	2.544	2.5044
12		22.548			16.365	

When  $a_n^2$  is plotted against  $n$ , the points lie nearly on a straight line. In Awbery's method (Searle, 'Experimental Physics,' p. 360), the "best" straight line passes through the centroids of the two groups of points. If the co-ordinates of the centroids of the first and second groups be  $h$ ,  $p$  and  $k$ ,  $q$ , we have, for air,  $h=10$ ,  $p=\frac{1}{3} \times 57.586=19.1953$ , and  $k=4$ ,  $q=\frac{1}{3} \times 22.548=7.5160$ . If the straight line be  $y=m_A x + c_A$ , where the axes of  $x$  and  $y$  are those of  $n$  and  $a^2$ , we have

$$m_A = (p - q) / (h - k) = \frac{1}{6} \times 11.6793 = 1.9466.$$

Then  $c_A = q - 4m_A = -0.2704$ . The values of  $a_n^2$  calculated, for  $x=12$ , 10, . . ., with these values of  $m_A$  and  $c_A$  are given in the Table. For water, we find  $m_W = 1.4753$ ,  $c_W = -0.4462$ . For the refractive index of water, we have  $\mu = m_A / m_W = 1.9466 / 1.4753 = 1.319$ .

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



LXXVI. *Plastic Stresses in a Semi-Infinite Cohesive Mass due to a Surcharge covering a Strip of the Surface of Infinite Length.*

By N. S. BOULTON, M.Sc., Assoc.M.Inst.C.E.,  
University of Sheffield\*.

[Received July 17, 1946.]

*Introduction.*

A SEMI-INFINITE fragmental mass having cohesion and internal friction, with a horizontal surface, is loaded on a strip of the surface of infinite length so as to produce a state of "plastic equilibrium."

The strip is assumed to be loaded uniformly along its length so that the state of plastic stress may be regarded as two-dimensional, being the same for all vertical planes perpendicular to the length of the strip. It is also assumed that the distribution of loading across the strip is symmetrical about the centre-line.

A rigorous formula for the bearing pressure on the strip was obtained by Prandtl<sup>(1)</sup> on the assumption that the mass is weightless. His theory showed that a uniform distribution of pressure across the strip is required to maintain plastic equilibrium. The theory was later extended by Reissner<sup>(2)</sup> to allow for a uniform vertical pressure applied to the surface of the mass outside the strip.

When allowance is made for the *weight* of the mass, both the shape of the slip-surfaces and the distribution of pressure across the loaded strip are modified. Reissner gave general equations which allow for the weight of the material, but a general solution of these equations, as applied to the problem considered in this paper, is as yet unknown.

In two limiting cases, however, rigorous solutions are available. For purely cohesive materials, having no angle of internal friction, the weight of the material has no effect on the load required to maintain plastic equilibrium. Also, for cohesionless materials, a solution can be found (provided the pressure  $\gamma h = 0$  (fig. 1)) by the methods developed by Ohde<sup>(3)</sup>, Jáky and others, though these methods involve rather laborious numerical or graphical calculations.

In the present paper a solution expressed as the first three terms of a series for the stress-function is developed by successive approximation. Over a certain limited range of the variables involved, the solution satisfies the general equations to a good approximation, and may therefore be considered as rigorous within this range for the conditions assumed.

---

\* Communicated by the Author.



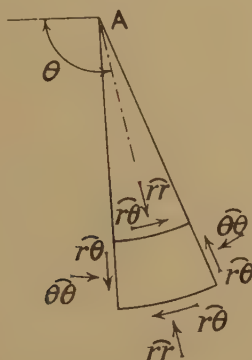
load, and has therefore developed internal frictional resistance ( $\phi$ ) as well as cohesion ( $c$ )\*.

### MATHEMATICAL SOLUTION.

*Solution for the Zone of Radial Shear (ACD in fig. 1).*

Polar co-ordinates are used with A as the origin of  $r$  and AB as the initial line from which  $\theta$  is measured in an anti-clockwise direction. Regarding compressive stress as positive the stress components are denoted by the symbols shown in fig. 2.

Fig. 2.



The usual equation is taken to represent the state of plastic equilibrium. It is assumed that the material at every point of the plastic region is about to fail by shear; that is, that a state of incipient plastic flow exists throughout the mass. The limiting shear stress on any slip-surface is assumed to be given by

$$s = c + \widehat{nn} \tan \phi,$$

where  $c$  and  $\phi$  are constants,  $c$  being the cohesion of the material,  $\phi$  the angle of internal friction and  $\widehat{nn}$  the normal pressure intensity on the slip-surface.

From the Mohr diagram shown in fig. 3, the equation of plastic equilibrium is at once obtained. For the radius of the Mohr-circle is given by

$$\sqrt{\left(\frac{\widehat{rr} - \widehat{\theta\theta}}{2}\right)^2 + \widehat{r\theta}^2}$$

and also by

$$c \cos \phi + \frac{\widehat{rr} + \widehat{\theta\theta}}{2} \sin \phi.$$

\* It is known that for soft clay  $\phi=0$  immediately after the application of the load. After consolidation  $\phi$  may be of the order of  $20^\circ$ . The practical significance of the ultimate bearing capacity based on the latter value of  $\phi$  is that it enables the ultimate factor of safety of the footing to be estimated.





Now evaluate the squares and neglect terms containing powers of  $\rho$  higher than the second; it is found that

$$2p_0(p_1+p_2)+2s_0(s_1+s_2)+p_1^2+s_1^2=2\sqrt{p_0^2+s_0^2}(q_1+q_2)k+q_1^2k^2.$$

The terms containing each power of  $\rho$  may be separately equated. Thus equating terms involving  $\rho$ ,

$$p_0p_1+s_0s_1-kq_1\sqrt{p_0^2+s_0^2}=0, \quad \dots \quad (5)$$

and equating terms containing  $\rho^2$ ,

$$p_0p_2+s_0s_2+\frac{1}{2}(p_1^2+s_1^2)-kq_2\sqrt{p_0^2+s_0^2}-\frac{1}{2}q_1^2k^2=0. \quad \dots \quad (6)$$

Finally, eliminating  $q_1$  between (5) and (6),

$$p_0p_2+s_0s_2+\frac{1}{2}(p_1^2+s_1^2)-kq_2\sqrt{p_0^2+s_0^2}-\frac{(p_0p_1+s_0s_1)^2}{2(p_0^2+s_0^2)}=0. \quad \dots \quad (7)$$

Equations (4), (5) and (7) replace equation (1), the equation of plastic equilibrium, in the following analysis:—

To satisfy the stress equations of equilibrium, the stresses are expressed in terms of a stress function. Write

$$\chi = \frac{c'^3}{\gamma^2} \left\{ \frac{\rho^2}{2} f_0(\theta) + \frac{\rho^3}{6} f_1(\theta) + \frac{\rho^4}{12} f_2(\theta) \dots \right\}, \quad \dots \quad (8)$$

where  $f_0$ ,  $f_1$  and  $f_2$  are functions of  $\theta$  and  $\phi$  only.

The stress components are given by

$$\left. \begin{aligned} \widehat{\theta\theta} &= \frac{\gamma^2}{c'^2} \frac{\partial^2 \chi}{\partial \rho^2} + c' \rho \sin \theta, & \widehat{r\theta} &= -\frac{\gamma^2}{c'^2} \frac{\partial}{\partial \rho} \left( \frac{1}{\rho} \frac{\partial \chi}{\partial \theta} \right), \\ \widehat{r'r} &= \frac{\gamma^2}{c'^2} \left( \frac{1}{\rho} \frac{\partial \chi}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 \chi}{\partial \theta^2} \right) + c' \rho \sin \theta. \end{aligned} \right\} \quad \dots \quad (9)$$

From (2), (8) and (9) it is easily found that

$$p_0 = \frac{1}{4} f_0'', \quad q_0 = f_0 + \frac{1}{4} f_0'', \quad s_0 = -\frac{1}{2} f_0',$$

where ' denotes  $\partial/\partial\theta$ . Substituting these values into (4),

$$\frac{1}{2} \sqrt{(\frac{1}{4} f_0''^2 + f_0'^2)} - k(f_0 + \frac{1}{4} f_0'') = \lambda \cos \phi. \quad \dots \quad (10)$$

The solution to this equation, which was obtained by Prandtl, is

$$f_0 = -\lambda/\mu + A e^{-2\mu\theta}, \quad \dots \quad (11)$$

where  $A$  is a constant.

Differentiating (11) with respect to  $\theta$ ,

$$f_0' = -2\mu A e^{-2\mu\theta}, \quad f_0'' = 4\mu^2 A e^{-2\mu\theta}, \quad \dots \quad (12)$$

and therefore

$$\left. \begin{aligned} p_0 &= \mu^2 A e^{-2\mu\theta}, & s_0 &= \mu A e^{-2\mu\theta}, \\ \sqrt{(p_0^2 + s_0^2)} &= \mu |A| e^{-2\mu\theta} \sqrt{(1 + \mu^2)}. \end{aligned} \right\} \quad \dots \quad (13)$$

Again, from (2), (8) and (9),

$$\left. \begin{aligned} p_1 &= \frac{1}{4} \rho \left( \frac{f_1''}{3} - f_1 \right), & s_1 &= -\frac{1}{3} \rho f_1', \\ q_1 &= \frac{1}{4} \rho \left( \frac{f_1''}{3} + 3f_1 \right) + \rho \sin \theta. \end{aligned} \right\} \quad \dots \quad (13a)$$

Substitute these values into (5). Using (13) and (13a), it is found after some simplification that

$$f_1' + 3\mu f_1 + 3\mu \sin \theta = 0. \quad (14)$$

This is a linear differential equation of the first order, the solution to which is

$$f_1(\theta) = \frac{3\mu}{1+9\mu^2} (\cos \theta - 3\mu \sin \theta) + B e^{-3\mu(\theta-\alpha)}, \quad (15)$$

where  $B$  is an arbitrary constant. The values of  $A$  and  $B$  are determined so as to make the stresses continuous at AD (fig. 1).

Thus, considering the stresses at the slip-plane AD which divides the zone of radial shear from the Rankine zone of passive earth pressure ADF\* (fig. 1), the normal and tangential stress components on AD for the zone ADF are readily found from the Mohr diagram (fig. 3). The minor principal stress, acting vertically, is  $\gamma(h+r \cos \alpha)$ . The normal stress

$$\begin{aligned} \widehat{\theta\theta} &= c \cos \phi + \gamma(1+k)(h+r \cos \alpha) \\ &= c' \left\{ \lambda \cos \phi + (1+k) \left( \frac{\gamma h}{c'} + \rho \cos \alpha \right) \right\}, \end{aligned} \quad (16)$$

and the shear stress  $\widehat{r\theta} = c + \mu \widehat{\theta\theta}$ .

Now, considering the part of  $\widehat{\theta\theta}$  which does not involve  $\rho$ , putting

$$\theta = \frac{\pi}{2} + \alpha \text{ in (11), } f_0(\tfrac{1}{2}\pi + \alpha) = -\lambda/\mu + A e^{-2\mu(\frac{1}{2}\pi + \alpha)},$$

which is the value of  $\widehat{\theta\theta}/c'$  at point A (fig. 1). Equating this value to the part independent of  $\rho$  in (16) and solving for  $A$ , gives

$$A = \frac{1+k}{\mu} e^{2\mu(\frac{1}{2}\pi + \alpha)}.$$

Substituting into (11),

$$f_0(\theta) = -\lambda/\mu + F_0(\theta),$$

where

$$F_0(\theta) = \frac{1+k}{\mu} e^{2\mu(\frac{1}{2}\pi + \alpha - \theta)}. \quad (17)$$

Similarly, the constant  $B$  in (15) is found by equating the coefficient of  $\rho$  in the value of  $\widehat{\theta\theta}$  at the plane AD for the radial shear zone to the corresponding value in (16).

$$\text{Thus } \frac{-3\mu}{1+9\mu^2} (\sin \alpha + 3\mu \cos \alpha) + B e^{-1\mu\pi} + \cos \alpha = (1+k) \cos \alpha.$$

$$\text{Whence } B = e^{\frac{1}{2}\mu\pi} \left\{ k \cos \alpha + \frac{3\mu}{1+9\mu^2} (\sin \alpha + 3\mu \cos \alpha) \right\}, \quad (18)$$

which can be evaluated for a given value of  $\phi$ .

---

\* The state of stress and shear pattern for this zone are described by Terzaghi : Ref. 4 Art. 12 pp. 35-38.

The value of  $f_2$  is obtained in a similar manner to  $f_1$ . From (2), (8) and (9) it is found that

$$\left. \begin{aligned} p_2 &= \frac{\rho^2}{6} \left( \frac{f_2''}{4} - 2f_2 \right), & s_2 &= -\frac{\rho^2}{4} f_2', \\ q_2 &= \frac{\rho^2}{6} \left( \frac{f_2''}{4} + 4f_2 \right). \end{aligned} \right\} \dots \dots (19)$$

Substituting for  $p_0$  and  $s_0$  from (13) into (7);

$$\mu p_2 + s_2 - \mu q_2 = -\frac{e^{2\mu\theta}(p_1 - \mu s_1)^2}{2\mu(1 + \mu^2)A} \dots \dots (20)$$

Now substitute for  $p_2$ ,  $s_2$  and  $q_2$  from (19) into (20), and also for  $p_1$  and  $s_1$  from (13 a) into (20). After some simplification it is found that

$$f_2' + 4\mu f_2 = \frac{e^{2\mu\theta}(f_1'' + 4\mu f_1' - 3f_1)^2}{72\mu(1 + \mu^2)A} \dots \dots (21)$$

Since  $f_1$  is known from (15) and (18) and  $A$  has also been determined, the right-hand side of (21) is a known function of  $\theta$ . A particular integral of (21) is found without difficulty, though the algebra is a little long. Adding the complementary function, involving an arbitrary constant  $D$ , the complete solution is

$$f_2(\theta) = \frac{e^{-2\mu(\frac{1}{2}\pi + \alpha)}}{(1 + \mu^2)(1 + k)} \{ e^{-\mu\theta}(A_1 \sin \theta + A_2 \cos \theta) + e^{2\mu\theta}(A_3 \sin 2\theta + A_4 \cos 2\theta) + A_5 e^{2\mu\theta} + A_6 \theta e^{-4\mu\theta} + D e^{-4\mu\theta} \}, \dots \dots (22)$$

where

$$\begin{aligned} A_1 &= \frac{B\mu(1 + \mu^2)(1 - 3\mu^2)}{(1 + 9\mu^2)^2} e^{3\mu\alpha}, & A_2 &= A_1 \frac{\mu(5 + 9\mu^2)}{1 - 3\mu^2}, \\ A_3 &= \frac{\mu^2(1 - 10\mu^2 - 27\mu^4)}{2(1 + 9\mu^2)^3}, & A_4 &= \frac{\mu^3(7 + 18\mu^2 + 27\mu^4)}{2(1 + 9\mu^2)^3}, \\ A_5 &= \frac{\mu(1 + \mu^2)}{6(1 + 9\mu^2)}, & A_6 &= \frac{B^2}{8} (1 + \mu^2)^2 e^{6\mu\alpha}. \end{aligned}$$

Equation (16), giving  $\widehat{\theta\theta}$  for the plane AD (fig. 1) of the Rankine zone, does not contain  $\rho^2$ . Therefore the coefficient of  $\rho^2$  in  $\widehat{\theta\theta}$  for the radial shear zone must vanish when  $\theta = \frac{1}{2}\pi + \alpha$ . Hence the constant  $D$  in (22) is found from the relation  $f_2(\frac{1}{2}\pi + \alpha) = 0$ .

#### *Continuity of Stresses at the Plane AD (fig. 1).*

The normal stresses on the plane AD have been made continuous by determining the constants  $A$ ,  $B$  and  $D$  so that values of  $\widehat{\theta\theta}$  found from the passive earth pressure zone agree with those from the zone of radial shear. For a correct solution it is also necessary that the shear stresses on AD should be continuous. This is true for the Prandtl solution represented by  $f_0$  in equation (8), since AD is a slip plane in both zones and therefore  $\widehat{r\theta} = c + \mu\widehat{\theta\theta}$ . The continuity of  $\widehat{r\theta}$  thus follows from that of  $\widehat{\theta\theta}$ .

Now, considering the stresses due to  $f_1$  in equation (8),

$$\widehat{r\theta} = -\frac{c'}{3}\rho f_1' \quad \text{and} \quad \widehat{\theta\theta} = c'\rho(f_1 + \sin \theta).$$

On substituting for  $f_1$  and  $f_1'$  from these equations into (14),  $\widehat{r\theta} = \mu\widehat{\theta\theta}$  for any radial plane. Hence the normal and shear stresses due to  $f_1$  are continuous at AD.

Again, if values of  $f_1''$ ,  $f_1'$  and  $f_1$ , obtained from (15) and (18), are substituted into (21), it is found that the expression on the right-hand side of (21) vanishes when  $\theta = \frac{1}{2}\pi + \alpha$ . Thus

$$f_2' + 4\mu f_2 = 0 \text{ at the plane AD.} \quad (23)$$

Hence, since  $f_2 = 0$  at AD, therefore  $f_2' = 0$ . Also, from (8) and (9), the radial shear stress due to  $f_2$  is  $-c'\rho^2 f_2'/4$ , which therefore vanishes at the plane AD. But, from (16), the shear stress at AD for the zone ADF does not contain a term in  $\rho^2$ . Hence the shear stress due to  $f_2$  is continuous at AD.

It follows that both the normal and shear stresses, represented by the first three terms in equation (8), are continuous at the plane AD.

#### *Solution for the Central Wedge, ABC (fig. 1).*

The compressive stress normal to AC (fig. 1), obtained from equations (8) and (9) (putting  $\theta = \alpha$ ), is

$$\widehat{\theta\theta} = c'[f_0(\alpha) + \rho\{f_1(\alpha) + \sin \alpha\} + \rho^2 f_2(\alpha) \dots]. \quad (24)$$

A solution for the plastic stresses in the zone ABC, agreeing with (24) on the plane AC, will now be developed. It is convenient to use rectangular co-ordinates  $x$  and  $y$  referred to C as origin (fig. 1). The stress components are expressed in terms of the dimensionless quantities, proportional to the co-ordinates, defined by  $\xi = \gamma x/c'$ ,  $\eta = \gamma y/c'$ .

It is assumed (as when defining  $\rho$ ) that  $\gamma/c'$  is sufficiently small that terms in the expressions for the stresses of higher degree in  $\xi$  and  $\eta$  than the second may be neglected.

Write for shortness,

$$\left. \begin{aligned} (\widehat{yy} - \widehat{xx})/2c' &= p_0 + p_1 + p_2 \dots, \\ (\widehat{yy} + \widehat{xx})/2c' &= q_0 + q_1 + q_2 \dots, \\ \widehat{xy}/c' &= s_0 + s_1 + s_2 \dots, \end{aligned} \right\}, \quad (25)^*$$

the suffix now denoting the sum of the powers of  $\xi$  and  $\eta$  in the term.

Using the present co-ordinates, the equation of plastic flow is

$$\left(\frac{\widehat{yy} - \widehat{xx}}{2}\right)^2 + \widehat{xy}^2 = \left(c \cos \phi + \frac{\widehat{yy} + \widehat{xx}}{2} k\right)^2. \quad (26)$$

Substituting from (25) into (26),

$$(p_0 + p_1 + p_2 \dots)^2 + (s_0 + s_1 + s_2 \dots)^2 = \{\lambda \cos \phi + (q_0 + q_1 + q_2 \dots)k\}^2. \quad (27)$$

\* The symbols  $p$ ,  $q$ , and  $s$  now have different meanings from those which were assigned to them in equations (2).



Since the stress distribution is symmetrical about the centre-line CE, there is no shear stress on CE (fig. 1). Therefore,  $s_0$  (which is proportional to the vertical shear stress at C) is zero. Also, at C,  $x=y=0$  and (27) becomes  $|p_0|-q_0k=\lambda \cos \phi$ . Now,  $\widehat{yy} > \widehat{xx}$ . Therefore  $p_0$  is positive and  $|p_0|=p_0$ . Thus

$$p_0 - q_0 k = \lambda \cos \phi. \quad \dots \dots \dots (28)$$

Substituting from (28) into (27) and putting  $s_0=0$ ,

$$(p_0 + p_1 + p_2 \dots)^2 + (s_1 + s_2 \dots)^2 = \{p_0 + (q_1 + q_2)k\}^2. \quad \dots \dots (29)$$

Terms on either side of (29) of a given degree in  $\xi$  and  $\eta$  may be separately equated. Thus, equating terms of the first degree,

$$p_1 = q_1 k. \quad \dots \dots \dots (30)$$

Equating terms of the second degree,

$$2p_0 p_2 + p_1^2 + s_1^2 = 2p_0 q_2 k + q_1^2 k^2.$$

Substituting for  $q_1 k$  from (30) and simplifying

$$p_2 + \frac{s_1^2}{2p_0} = q_2 k. \quad \dots \dots \dots (31)$$

Now, when the weight of the material is neglected, the stress components are constant at all points of the central wedge. Therefore, since powers of the small quantities  $\xi$  and  $\eta$  which are higher than the second are neglected, the stress components may be represented by a quadratic function of  $\xi$  and  $\eta$ . It follows that the stress function will be an algebraic function of the fourth degree.

Therefore, assume  $\chi = \chi_0 + \chi_1 + \chi_2 \dots$ , where

$$\left. \begin{aligned} \chi_0 &= \frac{c'^3}{\gamma^2} (a_0 \xi^2 + b_0 \eta^2), \\ \chi_1 &= \frac{c'^3}{\gamma^2} (a_1 \xi^3 + b_1 \xi^2 \eta + c_1 \eta^3), \\ \chi_2 &= \frac{c'^3}{\gamma^2} (a_2 \xi^4 + b_2 \xi^3 \eta + c_2 \xi^2 \eta^2 + d_2 \eta^4). \end{aligned} \right\} \dots \dots \dots (32)$$

The stress components are given by

$$\left. \begin{aligned} \widehat{xx} &= \frac{\gamma^2}{c'^2} \frac{\partial^2 \chi}{\partial \eta^2} + c'(\eta_0 - \eta), \\ \widehat{yy} &= \frac{\gamma^2}{c'^2} \frac{\partial^2 \chi}{\partial \xi^2} + c'(\eta_0 - \eta), \\ \widehat{xy} &= -\frac{\gamma^2}{c'^2} \frac{\partial^2 \chi}{\partial \xi \partial \eta}, \end{aligned} \right\}, \quad \dots \dots \dots (33)$$

$\eta_0$  being the value of  $\eta$  at E (fig. 1).

Terms which would contribute a shear stress on CE (fig. 1) have been omitted from equations (32).

In the following work the algebra is straight-forward but rather lengthy, and therefore only an outline of the analysis is given.

From equations (25), (32) and (33) the values of  $p_0, p_1, p_2, q_0, q_1, q_2, s_1$  and  $s_2$  are obtained in terms of the coefficients  $a_0, b_0, a_1$ , etc. These values are substituted into equations (28), (30) and (31). Terms independent of  $\xi$  and  $\eta$  are separately equated. Also, the coefficients of  $\xi, \eta, \xi\eta, \xi^2$  and  $\eta^2$  are equated in turn. Thus it is found that  $a_1 = b_2 = 0$ . The six equations obtained enable six of the coefficients to be eliminated from equations (32) and, as a result, the only unknown coefficients in (32) are  $a_0$  in  $\chi_0, b_1$  in  $\chi_1$  and  $a_2, a_0$  in  $\chi_2$ . The expressions for  $\chi_0, \chi_1$  and  $\chi_2$  are then expressed in polar co-ordinates  $r'\theta'$  referred to C as the origin of  $r'$  and the  $x$ -axis as the origin of  $\theta'$ . Using the first equation of (9) and putting  $\theta' = \alpha$ , the value of the normal stress on AC (fig. 1) is obtained in terms of the coefficients  $a_0, b_1$  and  $a_2$ . This stress distribution has also been found from consideration of the radial shear zone ACD and is given by equation (24),  $\rho$  (and  $r$ ) there being referred to A as origin. To identify the two stress distributions, (24) is expressed in terms of  $\rho'$ , using the relation  $\rho = \beta \sec \alpha - \rho'$ , where  $\beta = \gamma b/c'$ ,  $b$  being the half-width of the loaded strip. Finally, in the two expressions thus found for the normal stress on AC, the coefficients of  $\rho', \rho'^2$ , and those not involving  $\rho'$  are separately equated and the values of  $a_0, b_1$  and  $a_2$  are thus determined. All the coefficients in  $\chi$  of (32) are then known from previous equations.

The total vertical force on the loaded strip is of most practical interest and the expression for this alone is given. Denoting the vertical pressure on AB (fig. 1) by  $P$ ,

$$P = 2 \int_0^b \bar{y}y \cdot dx = \frac{2c'}{\gamma} \int_0^\beta \bar{y}y \cdot d\xi.$$

Using the values of the coefficients in (32), obtained by the procedure above outlined, it is found that

$$P = \frac{2c'^2}{\gamma} \left\{ \left( n_0 + \frac{\gamma h}{c'} \right) \beta + n_1 \beta^2 - n_2 \beta^3 \right\}, \quad \dots \quad (34)^*$$

where

$$n_0 = \frac{1}{2} F_0(\alpha) \sec^2 \alpha - \frac{1}{\mu},$$

$$n_1 = \frac{1}{4} \sec^3 \alpha \{ f_1(\alpha) - \sin \alpha \cos 2\alpha \},$$

$$n_2 = -\frac{1}{6} f_2(\alpha) \sec^4 \alpha - \frac{2}{3} \frac{u_1^2(\alpha)}{(1-k)u_0(\alpha)},$$

$$u_0(\alpha) = \{ F_0(\alpha) + f_1(\alpha) \beta \sec \alpha + \beta^2 f_2(\alpha) \sec^2 \alpha + \beta \tan \alpha \} \cos^2 \alpha \cos 2\alpha,$$

$$u_1(\alpha) = \frac{1}{4} \{ f_1(\alpha) + 2\beta f_2(\alpha) \sec \alpha - \cos 2\alpha \sin \alpha \},$$

$$\alpha = \pi/4 + \phi/2, \quad \beta = \gamma b/c', \quad c' = c + \mu\gamma h, \quad \mu = \tan \phi, \quad k = \sin \phi.$$

The coefficients  $n_0, n_1, n_2$  are functions of the single quantity  $\phi$  and are evaluated without difficulty.

Equation (34) has been obtained so as to make the normal stress on the plane AC (fig. 1) continuous along any normal to AC. But, for a correct

\* For the meanings of  $F_0, f_1$  and  $f_2$ , see equations (17), (15) and (22).

solution, both the normal and shear components acting on the slip-surface separating the central and radial shear zones must be continuous. It can be verified (by rather long analysis not here included) that the above solution satisfies this requirement, to the second order of approximation, along the slip-curve passing through A (fig. 1) which separates the two zones. This slip-curve makes an angle with the centre line EC of  $(\frac{1}{2}\pi - \alpha)$  where it intersects EC.

*Calculation of the Total Pressure on the Loaded Strip required to maintain Plastic Equilibrium.*

The values of the functions  $F_0(\alpha)$ ,  $f_1(\alpha)$  and  $f_2(\alpha)$ , for various values of  $\phi$  between 0 and  $34^\circ$ , have been calculated. The corresponding values of the coefficients  $n_0$ ,  $n_1$  and  $n_2$  (defined in (34)) are given in Tables I. and II.

For  $\phi > 34^\circ$  the three terms included in equation (34) will only give a good approximation to  $P$  for small values of  $\beta$  outside the usual practical range. No numerical values for  $\phi > 34^\circ$  are therefore included.

TABLE I.

$\phi^\circ$	$n_1$	$n_2$		
		$\beta=0$	$\beta=0.2$	$\beta=0.4$
0	0.0	0.0	0.0	0.0
8	0.9907	—	—	—
15	3.283	0.82	0.73	0.65
25	14.327	6.9	5.4	4.6
34	57.4	49.0	34.6	29.0

TABLE II.

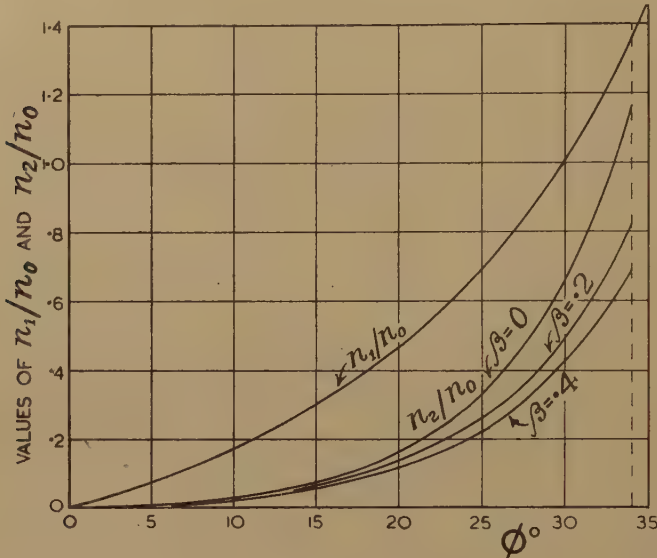
$\phi^\circ$	$n_0$	$\phi^\circ$	$n_0$	$\phi^\circ$	$n_0$	$\phi^\circ$	$n_0$
0	5.14	10	8.34	20	14.84	30	30.14
2	5.63	12	9.28	22	16.88	32	35.49
4	6.19	14	10.37	24	19.32	34	42.16
6	6.81	16	11.63	26	22.25	36	50.58
8	7.53	18	13.10	28	25.80	38	61.35

Values between those given in Table I. have been interpolated graphically and are represented by the curves in fig. 4. To give more convenient curves for plotting, the ratios  $n_1/n_0$  and  $n_2/n_0$  are plotted against  $\phi$ . When using these curves, a more suitable form of equation (34) is

$$P = \frac{2c'\beta}{\gamma} \left\{ \frac{\gamma h}{c'} + n_0 \left( 1 + \frac{n_1}{n_0} \beta - \frac{n_2}{n_0} \beta^2 \right) \right\}.$$

The retention of powers of  $\beta$  in  $u_0(\alpha)$  and  $u_1(\alpha)$  of equation (34), while going beyond the second order of approximation generally adopted in this paper, gives, it is believed, a more accurate result for  $P$ . In consequence,  $n_2/n_0$  varies slightly with  $\beta$  and therefore curves for different values of  $\beta$  are shown in fig. 4.

Fig. 4.



Values of  $P$  may easily be calculated using fig. 4 and values of  $n_0$  in Table II. For example, if the loaded strip is 8 ft. wide and rests on the surface of the soil,  $h=0$  and  $b=4$  ft. Assuming  $\phi=34^\circ$ ,  $\gamma=100$  lb./cu. ft.,  $c=1000$  lb./sq. ft., then  $\beta=0.4$ .

From fig. 4,  $n_1/n_0=1.361$ ,  $n_2/n_0=0.689$ . Substituting into equation (34),

$$\begin{aligned} P &= 337,300 + 183,700 - 37,200 \\ &= 483,800 \text{ lb.} \end{aligned}$$

It has not proved possible to find an upper limit for the error involved in equation (34), due to neglecting higher powers of  $\beta$ . For  $\phi=34^\circ$  and  $\beta=0.4$ , as in the previous example, the term involving  $n_2$  is about one-fifth of the previous term, and this gives some indication of the convergence in this case. For smaller values of  $\phi$  or  $\beta$  the convergence is more rapid.

#### Normal Stress Distribution on the Loaded Strip.

Putting  $\eta=\eta_0$  in the second of equations (33) and substituting for the derivative of  $\chi$  from (32),

$$\widehat{y}\eta=\eta_0=2c'(a_0+b_1\eta_0+c_2\eta_0^2+6a_2\xi^2), \quad \dots \quad (35)$$

which shows that the distribution of the normal stress on AB (fig. 1) is parabolic.



Assuming the numerical values adopted in the previous example, equation (35) becomes

$$\bar{y}y_{\eta=\eta_0}=62,241-33,074\xi^2. \quad \dots \quad (36)$$

The curve representing  $\bar{y}y$  is shown in fig. 5; it shows the stress decreasing towards the edges of the loaded strip, which is known to occur at the base of footings supported on soil having cohesion and internal friction when the plastic state is reached\*.

The total pressure per foot run of loaded area is easily found by integration of (36) to be  $P=483,800$  lb., in agreement with the previous value.

Fig. 5.

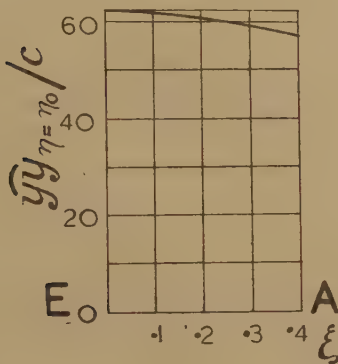
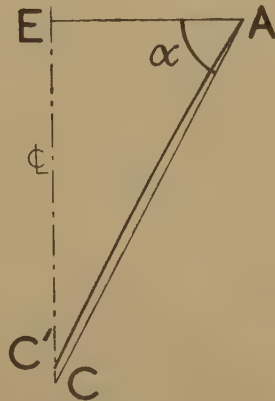


Fig. 6.



AC'.—Slip-curve separating zone of radial shear from central wedge.

#### *Comparison of the Value of $P$ with the Results of other Approximate Theories.*

To compare the total pressure  $P$ , as given by equation (34), with the approximate value given by the "Logarithmic spiral method" †, it is assumed that  $b=2$  ft. and that  $\gamma$ ,  $c$  and  $\phi$  have the same values as in the previous example. For a smooth foundation base, the logarithmic spiral method then gives  $P=212,300$  lb., while equation (34) gives  $P=209,000$  lb. The difference is 3300 lb., or 1.5 per cent., which is less than the maximum error of 3 per cent. associated with the logarithmic spiral method.

To avoid the rather laborious computation required in the spiral method, Terzaghi proposed the simpler linear form ‡

$$P=2b(cN_c+\gamma hN_q+\gamma bN_\gamma), \quad \dots \quad (37)$$

where  $N_c$ ,  $N_q$ , and  $N_\gamma$  are coefficients which depend only on the value

\* Terzaghi, Ref. 4, p. 391, fig. 126 c.

† Ref. 4, pp. 108-110.

‡ Ref. 4, p. 126, eq. (5).

of  $\phi$ . The  $N_\gamma$ -term represents the effect of the weight of the material on  $P$ , and Terzaghi determines its value by the spiral or friction-circle method, taking  $c=h=0$ .

The coefficients in equation (37) may be chosen so as to make this equation identical with (34), which is derived from the theory of the present paper. Then  $N_c=n_0$ ,  $N_q=\mu n_0+1$ ,  $N_\gamma=n_1-n_2\beta$ . Obtaining  $n_1$  and  $n_2$  from Table I. for  $\phi=34^\circ$  and  $\beta=0.2$ , it is found that  $N_\gamma=50.5$ . The logarithmic spiral method (taking  $c=0$ ) gives  $N_\gamma=45$ . The corresponding difference in  $P$ , as given by the two theories for the assumed values of the constants, is 2 per cent., which is well within the maximum error attributed by Terzaghi to the linear form (37) \*.

That equation (37), with Terzaghi's coefficients, agrees so closely with the more rigorous equation (34) is no doubt due to the fact that the slip-curve BCDF (fig. 1) differs but slightly, in the example considered, from its form when  $\gamma=0$ †.

For the numerical values assumed, there is thus close agreement, from a practical standpoint, between the ultimate bearing capacity as given by equation (34), the logarithmic spiral method, and equation (37).

It should be emphasized that though equation (34) appears complicated, its numerical evaluation, using Table II. and the curves in fig. 4, is as simple as the evaluation of Terzaghi's equation (37), and is much simpler to carry out than the logarithmic spiral method.

The slip-curve separating the zone of radial shear from the central wedge is shown as AC' in fig. 6. The equation to this curve has been deduced to the second order of approximation from the equations in the paper for  $\beta=0.2$  and the previously assumed values of the other constants. If the material were weightless the slip-curve would be the straight line AC inclined at an angle  $\alpha=62^\circ$  to the horizontal. The small departure of the curve from the straight line shows that the effect of the weight in modifying the slip-curve is small for the numerical values assumed.

#### References.

- (1) Prandtl, L., "Über die Härte plastischer Körper." *Nachr. kgl. Ges. Wiss. Göttingen, Math. phys. Klasse* (1920).
- (2) Reissner, H., "Zum Erddruckproblem," *Proc. 1st Intern. Congr. Applied Mechanics*, Delft (Holland) (1924).
- (3) Ohde, J., "Zur Theorie des Erddruckes unter besonderer Berücksichtigung der Erddruck Verteilung." *Die Bautechnik*, vol. xvi. pp. 241-245, 331-335 (1938).
- (4) Terzaghi, K., 'Theoretical Soil Mechanics,' p. 122. (Wiley, New York, 1943).

---

\* Ref. 4, p. 127.

† This is due to limiting  $\beta$  to small values to ensure accuracy in equation (34).

LXXVII. *A Method for the Measurement of the Dielectric Properties of Liquids in the Frequency Range 600–3,200 Mc./sec. (50–9.4 cm.).*

By R. DUNSMUIR, M.Sc. and J. G. POWLES, M.Sc.\*

Electrotechnics Department, Victoria University,  
Manchester.

[Received January 25, 1946.]

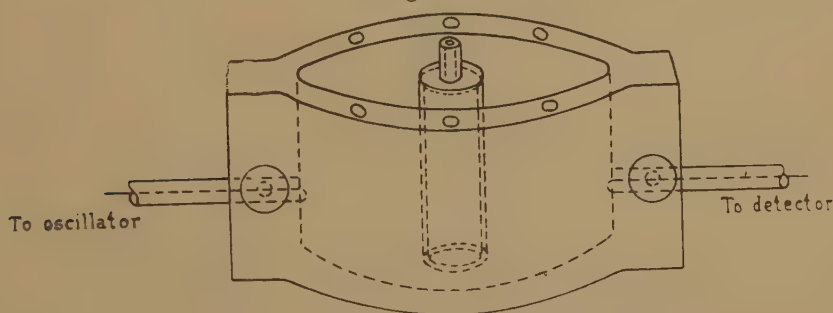
SUMMARY.

The paper describes a method of measuring the dielectric properties of liquids in the frequency range 600–3,200 Mc./sec. by modification of a technique for the measurement of solid dielectrics using a cylindrical cavity resonator<sup>(1)</sup>. The liquid is supported within the resonator in a thin-walled bottle, the effect of which is allowed for in the expressions for the permittivity and  $\tan \delta$  value of the liquid. By use of bottles of appropriate size measurements are possible over a wide range of permittivity and  $\tan \delta$  values.

Introduction.

A RECENT paper<sup>(1)</sup> describes the use of a circular cylindrical cavity resonator together with a variable frequency oscillator for measurement

\* Fig. 1.



$E_{010}$  Resonator containing bottle (lid removed).

of the permittivity and power factor ( $\tan \delta$ ) of cylindrical rod specimens of low loss solid dielectric materials at a frequency of 3,000 Mc./sec. The resonator is excited in the  $E_{010}$  mode of oscillation and the permittivity of the dielectric material is deduced from the resonant frequencies of the cavity when completely air-filled, and when containing the dielectric specimen. The dielectric power factor is obtained from the change in the corresponding quality ( $Q$ ) factors of the system. These  $Q$  factors are deduced from experimental resonance curves which relate the square

\* Communicated by Prof. Willis Jackson, D.Sc.

of the field strength within the resonator to the excitation frequency. With solid specimens accuracies of one per cent in permittivity and two per cent in power factor measurement can be attained.

When the method is applied to the measurement of liquid dielectrics, the latter must be retained in the necessary rod form by use of a thin-walled bottle, in the manner shown in fig. 1. An outline of the mathematical analysis for this three layer dielectric system, leading to expressions for the permittivity and  $\tan \delta$  value of the test liquid in terms of the resonant frequency and  $Q$  factor of the resonator with and without the specimen present, is given in section I. These expressions reduce to those derived for the two layer, solid dielectric case if the thickness of the bottle wall is reduced to zero. Section II. gives brief particulars of the experimental arrangements and the results of preliminary experiments carried out on a range of simple hydrocarbon liquids.

### THEORY OF THE $E_{010}$ RESONATOR CONTAINING THREE DIELECTRIC LAYERS.

#### *The Resonance Condition and the Permittivity Relations.*

The circular cylindrical resonator, shown diagrammatically in fig. 2,

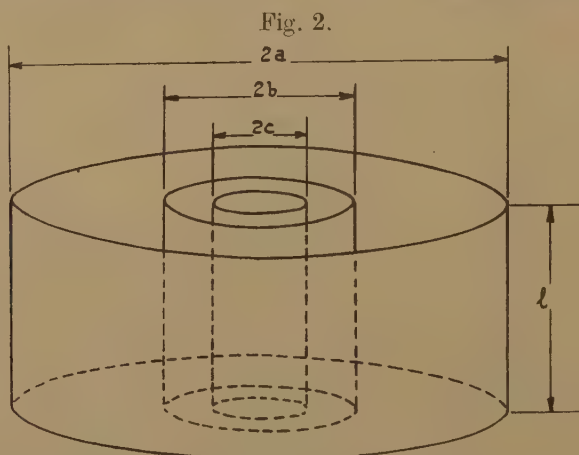


Fig. 2.

consists ideally of three regions denoted by subscripts (a), (b), (c) respectively, containing media of different dielectric properties. The boundary surfaces between adjacent regions are concentric cylinders of radii  $b$ ,  $c$  metres and the highly conducting outer wall is of radius  $a$  metres.

The field components for the  $E_{010}$  mode\* are then given in M.K.S. units and in terms of cylindrical polar co-ordinates by

$$\left. \begin{aligned} H_{\theta} &= \{A J_1(\beta r) + B Y_1(\beta r)\} e^{j\omega t} \text{ amps./metre,} \\ E_z &= -j \sqrt{\frac{\mu}{\epsilon}} \{A J_0(\beta r) + B Y_0(\beta r)\} e^{j\omega t} \text{ volts/metre,} \end{aligned} \right\} \quad \dots (1)$$

\* These are based on the justifiable assumption of a resonator wall of infinite conductivity.



where the amplitude constants  $A$ ,  $B$  and the propagation constant  $\beta = w\sqrt{\epsilon\mu}$  are chosen appropriately for each of the regions (a), (b), (c).

Relations between the values taken by  $A$ ,  $B$  within these regions are obtained from consideration of the physical conditions at the boundary surfaces, namely :

(1) that the tangential component of electric field intensity is zero at the outer wall  $r=a$ , and

(2) that the tangential components of electric and magnetic intensity are continuous at the boundary surfaces  $r=b$  and  $r=c$ .

From these are obtained the equations (2), in which it is assumed that the permeability has the value  $\mu_a$  everywhere within the system.

$$\left. \begin{aligned} A_a J_0(\beta_a a) + B_a Y_0(\beta_a a) &= 0, \\ A_a J_1(\beta_a b) + B_a Y_1(\beta_a b) - A_b J_1(\beta_b b) - B_b Y_1(\beta_b b) &= 0, \\ \frac{A_a}{\sqrt{\epsilon_a}} J_0(\beta_a b) + \frac{B_a}{\sqrt{\epsilon_a}} Y_0(\beta_a b) - \frac{A_b}{\sqrt{\epsilon_b}} J_0(\beta_b b) - \frac{B_b}{\sqrt{\epsilon_b}} Y_0(\beta_b b) &= 0, \\ A_b J_1(\beta_b c) + B_b Y_1(\beta_b c) - A_c J_1(\beta_c c) &= 0, \\ \frac{A_b}{\sqrt{\epsilon_b}} J_0(\beta_b c) + \frac{B_b}{\sqrt{\epsilon_b}} Y_0(\beta_b c) - \frac{A_c}{\sqrt{\epsilon_c}} J_0(\beta_c c) &= 0, \end{aligned} \right\} \dots (2)$$

where  $\beta_a = w\sqrt{\mu_a \epsilon_a}$ ,  $\beta_b = \sqrt{\frac{\epsilon_b}{\epsilon_a}} \beta_a$  and  $\beta_c = \sqrt{\frac{\epsilon_c}{\epsilon_a}} \beta_a$ .

Now define quantities  $M_1 \dots M_4$  and  $M$  as follows.

$$\left. \begin{aligned} M_1 &= J_0(\beta_b b) Y_1(\beta_b c) - J_1(\beta_b c) Y_0(\beta_b b), \\ M_2 &= J_1(\beta_b b) Y_1(\beta_b c) - J_1(\beta_b c) Y_1(\beta_b b), \\ M_3 &= J_0(\beta_b b) Y_0(\beta_b c) - J_0(\beta_b c) Y_0(\beta_b b), \\ M_4 &= J_1(\beta_b b) Y_0(\beta_b c) - J_0(\beta_b c) Y_1(\beta_b b), \end{aligned} \right\} \dots (3)$$

$$M = \frac{Y_1(\beta_a b)/J_1(\beta_a b) - Y_0(\beta_a a)/J_0(\beta_a a)}{Y_0(\beta_a b)/J_0(\beta_a b) - Y_0(\beta_a a)/J_0(\beta_a a)} = 1 + \frac{1}{F} \frac{a}{b} \frac{J_0(\beta_a a)}{J_1(\beta_a a)}, \dots (4)$$

where  $F = [Y_0(\beta_a a)J_0(\beta_a b) - Y_0(\beta_a b)J_0(\beta_a a)] \frac{\pi \beta_a a}{2} \dots (5)$

On elimination of the amplitude constants  $A_a$ ,  $A_b$ ,  $A_c$ ,  $B_a$  and  $B_b$  from equations (2), it follows that

$$\sqrt{\frac{\epsilon_c}{\epsilon_b}} \frac{J_1(\beta_a c)}{J_0(\beta_c c)} = \frac{M_2 - \sqrt{\frac{\epsilon_a}{\epsilon_b}} \frac{J_1(\beta_a b)}{J_0(\beta_a b)} M_1 M}{M_4 - \sqrt{\frac{\epsilon_a}{\epsilon_b}} \frac{J_1(\beta_a b)}{J_0(\beta_a b)} M_3 M} \dots (6)$$

The transcendental equation (6) relates the permittivities and the dimensions of the three regions to the resonant frequency, but is not

suitable for computation. It may be expressed in the more convenient form (10) by use of the approximation

$$\frac{J_1(x)}{J_0(x)} \doteq \frac{x}{2} \left[ 1 + \frac{x^2}{8} \right] \dots \dots \dots (7)$$

and the following approximate expressions for  $M_1 \dots M_4$

$$\left. \begin{aligned} M_1 &= -\frac{2}{\pi \beta_b c}, \\ M_2 &= M_1 \beta_b (b-c), \\ M_3 &= M_2 = M_1 \beta_b (b-c), \\ M_4 &= -\frac{c}{b} \cdot M_1. \end{aligned} \right\} \dots \dots \dots (8)$$

Expressions (8) are obtained by expansion in Taylor series assuming that  $(b-c)$  is small compared with  $b$  and  $c$  and by use of the identity <sup>(2)</sup>

$$J_1(x)Y_0(x) - Y_1(x)J_0(x) = \frac{2}{\pi x} \dots \dots \dots (9)$$

This leads to the equation

$$\frac{\epsilon_c}{\epsilon_a} = \frac{\frac{b^2}{c^2} \left[ 1 + \frac{(\beta_a b)^2}{8} \right] \left[ M - M_b \left( 1 - \frac{(\beta_a b)^2}{8} \right) \right]}{1 + \frac{(\beta_a b)^2}{8} [M - M_b]}, \dots \dots (10)$$

where

$$M_b = 2 \frac{\epsilon_b (b-c)}{\epsilon_a b}.$$

This expression (10) is suitable for direct calculation of the relative permittivity,  $\frac{\epsilon_c}{\epsilon_a}$ , of the test liquid. Computation of the function  $M$  equation (4), is assisted by use of the graphical plot of numerical values of the function  $F$ , equation (5), which is given in reference <sup>(1)</sup>.

The relative permittivity,  $\frac{\epsilon_b}{\epsilon_u}$  of the bottle material is supposed known but is not required to a high degree of accuracy. The degree of inaccuracy introduced by the approximations is less than one per cent. for values of  $\frac{\epsilon_b}{\epsilon_a}$  less than 4 and of  $\frac{b-c}{a}$  less than  $\frac{1}{40}$ .

In the case of very small bottles  $\left( \frac{b}{a} < \frac{1}{50} \right)$  further approximations lead to :—

$$\frac{\epsilon}{\epsilon_a} = 1 + 0.539 \frac{a^2 (df)}{c^2 f_0}, \dots \dots \dots (11)$$

where  $(df)$  is the change of resonant frequency on filling the bottle with liquid and  $f_0$  is the resonant frequency. This expression (11) is useful

for liquids of high loss ( $\tan \delta > 0.006$ ), in which case a specimen of small diameter must be used.

### Power Factor Relations.

Referring to the field equations (1), write for brevity

$$A_a J_0(\beta_a r) + B_a Y_0(\beta_a r) = {}_a Z_0(\beta_a r)$$

etc. where  $Z_0(x)$  is known as a cylinder function. The field equations for the region (a) then become

$$\left. \begin{aligned} H_\theta &= {}_a Z_1(\beta_a r) e^{j\omega t}, \\ E_z &= \sqrt{\frac{\mu_a}{\epsilon_a}} {}_a Z_0(\beta_a r) e^{j\omega t} \end{aligned} \right\} \dots \dots \dots (1')$$

and similar relations hold for the (b) and (c) regions.

The procedure for calculating the Q value of the system makes use of the relation

$$Q = w \frac{\text{stored energy}}{\text{mean power loss}} \dots \dots \dots (12)$$

and has been described in reference (1). Thus the energy stored is given by

$$W = 2\pi l \int_0^a r \frac{\mu H^2}{2} dr = \pi l \int_0^a r {}_a Z_1^2(\beta_a r) dr$$

where  $r$  embraces the regions (a), (b) and (c).

$$\begin{aligned} &= \frac{1}{2} l \pi \mu \left\{ a^2 {}_a Z_1^2(\beta_a a) + b^2 \left( 1 - \frac{\epsilon_a}{\epsilon_b} \right) {}_b Z_0^2(\beta_b b) \right. \\ &\quad \left. + c^2 \left( 1 - \frac{\epsilon_b}{\epsilon_c} \right) {}_c Z_0^2(\beta_c c) \right\} \dots \dots \dots (13) \end{aligned}$$

The power loss in the liquid dielectric (region (c)) is

$$\left. \begin{aligned} P &= 2\pi l \int_0^c r \frac{g_c E^2}{2} dr \\ &= \frac{1}{2} \pi l w \mu \tan \delta_c \cdot c^2 \{ {}_c Z_1^2(\beta_c c) + {}_c Z_0^2(\beta_c c) \} \end{aligned} \right\} \dots \dots \dots (14)$$

where  $g_c$  is the effective conductivity of the liquid and  $\tan \delta_c = g_c / w \epsilon_c$ .

Now define  $Q_2$  as the Q factor of the system with the bottle filled with test dielectric and  $Q_1$  as the corresponding value if the bottle were filled with loss-free liquid of the same permittivity. It then follows from (12) and (13) and the definition of Q that

$$\begin{aligned} \tan \delta_c &= \frac{a^2 {}_a Z_1^2(\beta_a a) + b^2 \left( 1 - \frac{\epsilon_a}{\epsilon_b} \right) {}_b Z_0^2(\beta_b b) + c^2 \left( 1 - \frac{\epsilon_b}{\epsilon_c} \right) {}_c Z_0^2(\beta_c c)}{c^2 \{ {}_c Z_1^2(\beta_c c) + {}_c Z_0^2(\beta_c c) \}} \\ &\quad \times \left[ \frac{1}{Q_2} - \frac{1}{Q_1} \right], \dots \dots \dots (15) \end{aligned}$$

which reduces to the approximate form (16) when the bottle wall is thin.

$$\tan \delta_c = \frac{\left\{ \frac{a^2}{c^2 F_b^2} + \frac{b^2}{c^2} \left( \frac{\epsilon_b}{\epsilon_a} - 1 \right) \right\} \left\{ 1 - 2\beta_c(b-c) \frac{J_1(\beta_c c)}{J_0(\beta_c c)} \right\} + \frac{\epsilon_c - \epsilon_b}{\epsilon_a}}{\frac{\epsilon_c}{\epsilon_a} \left\{ 1 + \frac{J_1^2(\beta_c c)}{J_0^2(\beta_c c)} \right\}} \times \left[ \frac{1}{Q_2} - \frac{1}{Q_1} \right] \quad (16)$$

in which  $F_b$  is the function  $F$ , defined in (5), expressed in terms of the ratio  $b/a$ .

When the relative permittivity  $\frac{\epsilon_b}{\epsilon_a}$  and the power factor  $\tan \delta_b$  of the bottle material are known (these are most conveniently obtained from measurements on a rod specimen of the material),  $Q_1$  may be derived from the  $Q$  factor  $Q_0$ , of the completely air-filled resonator by means of the relation

$$Q_1 = Q_0 \times \left\{ \frac{\left\{ \frac{a^2}{c^2 F_b^2} + \frac{b^2}{c^2} \left( \frac{\epsilon_b}{\epsilon_a} - 1 \right) \right\} \left\{ 1 - 2\beta_c(b-c) \frac{J_1(\beta_c c)}{J_0(\beta_c c)} \right\} + \frac{\epsilon_c - \epsilon_b}{\epsilon_a}}{\frac{a}{a+l} \sqrt{\frac{f_0}{f_1}} \left[ \left\{ \frac{a(a+l)}{l^2 F_b^2} + \frac{b^2}{c^2} \left( \frac{\epsilon_b}{\epsilon_a} - 1 \right) \right\} \left\{ 1 - 2\beta_b(b-c) \frac{J_1(\beta_c c)}{J_0(\beta_c c)} \right\} + \frac{\epsilon_c - \epsilon_b}{\epsilon_a} \right] + 2 \frac{b-c}{c} \frac{\epsilon_b}{\epsilon_a} Q_0 \tan \delta_b} \right\} \quad (17)$$

where the frequencies  $f_0$  and  $f_1$ , correspond to  $Q_0$  and  $Q_1$  respectively.

This complicated relation is not so formidable as at first appears since many of the terms have been computed already for use in (16). The two expressions (16) and (17) are accurate to better than one per cent for values of  $\frac{\epsilon_b}{\epsilon_a}$  less than 4, and of  $\frac{b-c}{a}$  less than  $1/40$ , as in the case of the permittivity relation. For bottles of small radius, with  $\frac{b}{a}$  less than  $1/50$ , expression (16) reduces to

$$\tan \delta = \frac{0.269}{\epsilon_c/\epsilon_a} \frac{a^2}{c^2} \left[ \frac{1}{Q_2} - \frac{1}{Q_1} \right] \quad (18)$$

## II. DESCRIPTION OF APPARATUS AND DEMONSTRATION OF THE ACCURACY OF THE METHOD.

Measurements have been carried out using four separate resonators, details of which are given in Table I.

At the two shorter wave-lengths the oscillators employed were of the reflection type and the coupling and calibrating arrangements were as described in reference (1). The longer wave-length resonators were energized from a coaxial line type triode oscillator provided with a



micrometer driven plunger for fine tuning. In this case frequency calibration was carried out with a coaxial line wave meter.

The test bottles were constructed with a neck penetrating through a hole drilled centrally in the resonator lid. Experiment showed that the presence of this hole exercised a negligible effect on the resonant frequency and *Q* factor.

Choice of material for the bottles was determined by the following requirements. It was necessary for the material to be readily workable

TABLE I.

Resonator diameter in cm.	Axial length of resonator in cm.	Resonant frequency air-filled in Mc./sec.	Free space wave-length of energizing oscillation at resonance, in cm.
6.58	3.77	3,490	8.60
7.15	3.77	3,210	9.35
17.78	10.2	1,290	23.3
34.66	15.2	663	45.2

TABLE II.

Bottle material	Bottle diameter in cm.	Wall thickness in cm.	Resonant diameter in cm.	Resonant wave-length air-filled in cm.	Measured relative permittivity of benzene
Polythene	1.110	0.060	6.58	8.60	2.31
"			7.15	9.35	2.28
"	1.112	0.077	6.58	8.60	2.24
"			7.15	9.35	2.30
Quartz	0.920	0.066	6.58	8.60	2.27
"			7.15	9.35	2.24
"	3.22	0.120	17.78	23.3	2.28
"	5.0	0.534	34.66	45.2	2.29

into the required shape, to be unaffected either physically or chemically by the test liquids, and to be of low permittivity and power factor in order to minimize the effect of the bottle on the resonance frequency and *Q* factor of the system.

Polythene naturally suggested itself and successful measurements of the permittivity of a range of liquids have been made using bottles of this material. Unfortunately, however, it was found that many organic liquids are absorbed by the polythene, distorting the bottle and producing conditions which prevented accurate delineation of resonance curves, and invalidated the use of the theoretical expressions.

Since glass of sufficiently low loss was not obtainable, bottles of transparent quartz (relative permittivity 3.7 and  $\tan \delta = 0.0001_5$ ) were constructed and found to be very satisfactory.

The consistency of the method and the validity of the theoretical analysis were checked by a series of measurements. The permittivity of the non-polar liquid benzene was measured using various resonators and bottles of different wall thickness and diameter, the results being given in Table II.

In Table III, permittivity values obtained at 3,000 Mc./sec. for a series of non-polar organic liquids are compared with the accepted values<sup>(3)</sup> at audio and optical frequencies.

TABLE III.

Liquid	n-Hexane	Cyclo-hexane	Carbon tetra-chloride	Benzene	Carbon bi-sulphide
Relative permittivity at 3,000 Mc./sec. . .	1.91	2.04	2.24	2.28	2.64
Accepted audio frequency values.	1.87	2.05	2.24	2.28	2.64
Square of refractive index $\eta_D^2$ . .	1.90	2.04	2.15	2.27	2.66

TABLE IV.

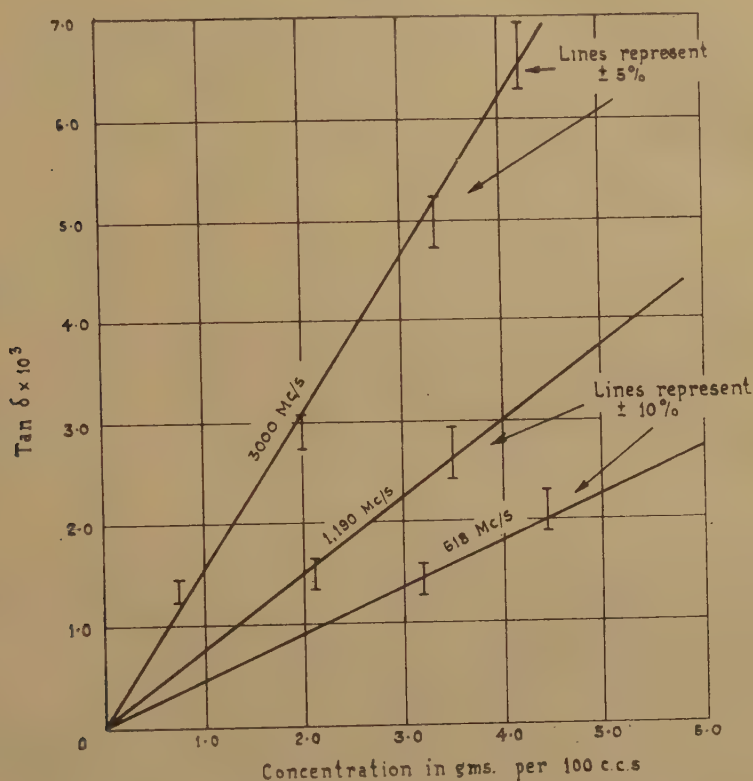
Bore of tube in mm.	Relative permittivity	$\tan \delta \times 10^3$
0.556	2.40	8.0
0.819	2.41	8.0
1.19	2.45	8.9
1.59	2.39	8.5

Fig. 3 shows the variation of  $\tan \delta$  with concentration for a solution of brombenzene in benzene at three of the four frequencies of measurement. In these measurements the loss contribution due to the solvent itself was allowed for in each case by substituting for  $Q_1$  in equation (16) the measured  $Q$  factor of the system when the bottle was filled with pure solvent.

The estimated accuracy of  $\tan \delta$  measurement at the two higher frequencies was  $\pm 5$  per cent and at the two lower frequencies  $\pm 10$  per cent. Greatest accuracy of  $\tan \delta$  measurement was achieved when introduction of the liquid decreased the  $Q$  value by a factor of about one-third. For lossy liquids it was necessary therefore to use a very small bottle.

In this case the loss in the bottle wall was less important and successful measurements have been made using fine capillary glass tubing; the expressions (11) and (18) were sufficiently accurate for these calculations.

Fig. 3.



Loss factor of solution of brombenzene in benzene at 19° c.  
Effect of concentration and frequency.

Table IV. gives results obtained in this way for pure toluene at 3,000 Mc./sec. using capillaries of various diameters.

#### Conclusion.

By use of the method described above and of an  $H_{01}$  resonator system permitting measurements at frequencies in excess of 3,200 Mc./sec., measurements have been made on a range of solutions of polar molecules in benzene. The results of these measurements and a discussion of them in terms of the theory of dilute solutions of polar molecules in non-polar solvents will form the subject matter of a later paper <sup>(4)</sup>.

#### Acknowledgment.

The work described formed part of a programme of extra-mural research carried out on behalf of the Ministry of Supply in the Electrotechnics

Dept., Manchester University, under the direction of Prof. Willis Jackson. The author's thanks are due to him, and also to the chief Scientific Officer of the Ministry for permission to publish.

### References.

- (1) "Resonance Methods of Dielectric Measurement at Centimetre Wavelengths," F. Horner, T. A. Taylor, R. Dunsmuir, J. Lamb and Willis Jackson, *Journal Inst. of Electrical Engineers*, xciii. part III. p. 53 (1946).
- (2) McLachlan, 'Bessel Functions for Engineers,' p. 156. (Oxford University Press.)
- (3) International Critical Tables.
- (4) Jackson & Powles, Faraday Society, Symposium on 'Dielectrics,' April 1946 (awaiting publication).

## LXXVIII. *Large Deflections of Circular and Square Plates.*

By H. D. CONWAY, M.A., Ph.D.\*

[Received August 29, 1945,]

### 1. *Introduction.*

FORMULÆ for the approximate determination of the deflections and stresses in plates are easily obtained from the principal of virtual displacements. This principle has been used by Timoshenko ('Theory of Plates and Shells,' Chapter IX.) for the solution of the problem of the uniformly loaded circular plate with a clamped edge. In the following analysis, formulæ are obtained for the deflections and stresses in the practical problems of uniformly loaded circular and square plates supported at the edges, the latter being sufficiently stiff to prevent displacements occurring there. By neglecting the tension terms, the equations are applicable to simply supported plates with small deflections and by neglecting the bending terms they are applicable to membranes.

### 2. *List of Symbols.*

- $u, v, w$  = displacement of point  $P(x, y, z)$  in directions  $Ox, Oy, Oz$ .
- $w_0$  = maximum deflection in direction  $Oz$ .
- $e_r$  = radial strain.
- $e_t$  = tangential strain.
- $U_1$  = strain energy due to bending.
- $U_2$  = strain energy due to tension.
- $U$  = total strain energy =  $U_1 + U_2$ .
- $p$  = normal pressure on plate surface.
- $M$  = bending moment per unit length of plate.

\* Communicated by the Author.



$f$ =bending stress.

$t$ =tensile stress.

$q$ =resultant stress.

$2a$ =plate diameter, plate side.

$h$ =plate thickness.

$E$ =Young's modulus.

$\sigma$ =Poisson's ratio

$D$ =flexural rigidity= $\frac{Eh^3}{12(1-\sigma^2)}$ .

### 3. Analysis.

#### (a) Circular plate.

Assume that the deflected form of the plate may be represented by the equation

$$w=w_0\left(1+m\frac{r^2}{a^2}+n\frac{r^4}{a^4}\right), \quad \dots \quad (1)$$

where  $m=-2(3+\sigma)/5+\sigma$  and  $n=(1+\sigma)/5+\sigma$ . This is the exact equation if stretching of the middle surface of the plate is neglected.

It is shown in books on elasticity that the strain energy of bending is given by

$$U_1=\frac{1}{2}D\int_0^{2\pi}\int_0^a\left[\left(\frac{\partial^2 w}{\partial r^2}\right)^2+\frac{1}{r}\frac{\partial w}{\partial r}\left(\frac{\partial w}{\partial r}+2\sigma\frac{\partial^2 w}{\partial r^2}\right)\right]\dots \quad (2)$$

Substituting equation (1) in equation (2), we have

$$U_1=\frac{32\pi D(1+\sigma)(7+\sigma)}{3(5+\sigma)^2}\cdot\frac{w_0^2}{a^2}\dots \quad (3)$$

For  $\sigma=0.25$ , equation (3) becomes

$$U_1=3.506\pi D\frac{w_0^2}{a^2}, \quad \dots \quad (4a)$$

or for  $\sigma=0.30$

$$U_1=3.602\pi D\frac{w_0^2}{a^2}\dots \quad (4b)$$

It has been shown (*e. g.* Timoshenko, 'Theory of Plates and Shells,' p. 334) that the strain energy due to stretching of a plate may be represented by

$$U_2=\frac{12\pi D}{h^2}\int_0^a(e_r^2+e_t^2+2\sigma e_r e_t)r\,dr, \quad \dots \quad (5)$$

where

$$e_r=\frac{du}{dr}+\frac{1}{2}\left(\frac{dw}{dr}\right)^2 \quad \text{and} \quad e_t=\frac{u}{r}.$$

We represent the radial displacement of the plate by the approximation

$$u=r(A+Br)(a-r), \quad \dots \quad (6)$$

from which it is seen that the centre and edges of the plate are not displaced radially.

Substituting equations (1) and (6) in (5), we have

$$\begin{aligned}
 U_2 = & \frac{12\pi D}{h^2} \left[ \frac{A^2 a^4}{4} + \frac{7B^2 a^6}{60} + \frac{3ABa^5}{10} \right. \\
 & - 4w_0^2 A \left\{ \frac{m^2 a}{20} (3-\sigma) + \frac{2mna}{21} (5-\sigma) + \frac{n^2 a}{18} (7-\sigma) \right\} \\
 & - 4w_0^2 B \left\{ \frac{m^2 a^2}{20} (3-\sigma) + \frac{mna}{14} (5-\sigma) + \frac{2n^2 a^2}{45} (7-\sigma) \right\} \\
 & \left. + \frac{4w_0^4}{a^2} \left( \frac{m^4}{6} + m^3 n + \frac{12}{5} m^2 n^2 + \frac{8}{3} mn^3 + \frac{8}{7} n^4 \right) \right] \dots \dots \dots (7)
 \end{aligned}$$

In the equilibrium position, the value of  $U_2$  will be a minimum and hence

$$\frac{\partial U_2}{\partial A} = \frac{\partial U_2}{\partial B} = 0. \dots \dots \dots (8)$$

We thus have two equations for the determination of A and B

$$\left. \begin{aligned}
 A + \frac{3}{5} Ba &= \frac{4w_0^2}{a^3} \left[ \frac{m^2}{10} (3-\sigma) + \frac{4mn}{21} (5-\sigma) + \frac{n^2}{9} (7-\sigma) \right], \\
 A + \frac{7}{9} Ba &= \frac{4w_0^2}{a^3} \left[ \frac{m^2}{9} (3-\sigma) + \frac{5mn}{21} (5-\sigma) + \frac{4n^2}{27} (7-\sigma) \right]
 \end{aligned} \right\} \dots \dots \dots (9)$$

Solving these equations, we obtain

$$\left. \begin{aligned}
 A &= \frac{w_0^2}{a^3} \left[ \frac{m^2}{4} (3-\sigma) + \frac{5mn}{21} (5-\sigma) - \frac{n^2}{18} (7-\sigma) \right], \\
 B &= \frac{w_0^2}{a^4} \left[ \frac{m^2}{4} (3-\sigma) + \frac{15mn}{14} (5-\sigma) + \frac{5n^2}{6} (7-\sigma) \right]
 \end{aligned} \right\} \dots \dots \dots (10)$$

For  $\sigma=0.25$

$$A = 0.8662 w_0^2/a^3, \quad B = -0.1268 w_0^2/a^4, \quad \dots \dots \dots (11a)$$

and, for  $\sigma=0.3$

$$A = 0.8537 w_0^2/a^3, \quad B = -0.1551 w_0^2/a^4. \quad \dots \dots \dots (11b)$$

Substituting in equation (7), we have for  $\sigma=0.25$

$$U_2 = 3.190\pi D \frac{w_0^4}{a^2 h^2}, \quad \dots \dots \dots (12a)$$

and, for  $\sigma=0.3$

$$U_2 = 3.362\pi D \frac{w_0^4}{a^2 h^2}. \quad \dots \dots \dots (12b)$$

Summing equations (4) and (12), the total strain energy U, for  $\sigma=0.25$  is given by

$$U = U_1 + U_2 = \pi D \frac{w_0^2}{a^2} \left( 3.506 + 3.190 \frac{w_0^2}{h^2} \right), \quad \dots \dots \dots (13a)$$

and, for  $\sigma=0.3$

$$U = U_1 + U_2 = \pi D \frac{w_0^2}{a^2} \left( 3.602 + 3.362 \frac{w_0^2}{h^2} \right). \quad \dots \dots \dots (13b)$$

From the principle of virtual displacements

$$\frac{dU}{dw_0} = 2\pi p \int_0^a \left(1 + m \frac{r^2}{a^2} + n \frac{r^4}{a^4}\right) r dr, \quad \dots \quad (14)$$

whence, for  $\sigma=0.25$

$$w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D} \frac{1}{1+1.820 \frac{w_0^2}{h^2}} \quad \dots \quad (15)$$

From this cubic equation the maximum deflection  $w_0$  can easily be calculated. It will be seen that for small values of  $w_0/h$ , we have

$$w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D}, \quad \dots \quad (16)$$

which is the equation obtained from the simple theory in which stretching is neglected.

For large values of  $w_0/h$ , we have for  $\sigma=0.25$

$$w_0 = 0.74 \sqrt[3]{\frac{pa^4}{Eh}} \quad \dots \quad (17)$$

as the equation for the deflection of a membrane. A more complete solution for the deflection of a membrane (Hencky, *Zeit. Math. Physik.* vol. lxiii. p. 311) gives a factor of 0.66 in equation (17).

For  $\sigma=0.3$ , equation (15) becomes

$$w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D} \frac{1}{1+1.867 \frac{w_0^2}{h^2}}, \quad \dots \quad (18)$$

which for large values of  $w_0/h$  gives

$$w_0 = 0.72 \sqrt[3]{\frac{pa^4}{Eh}} \quad \dots \quad (19)$$

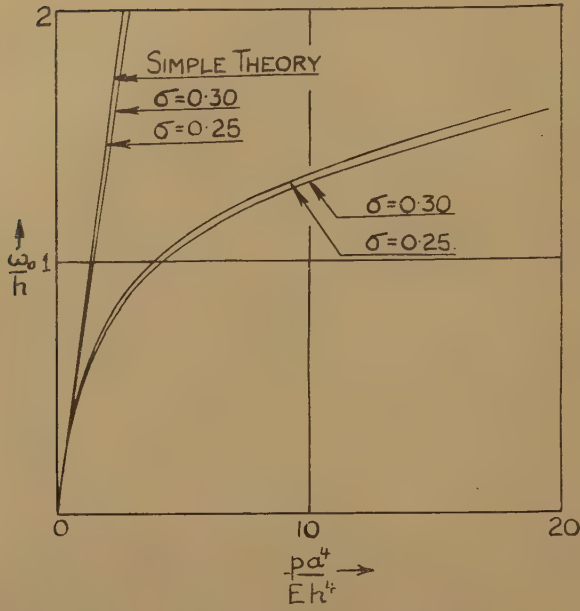
Graphs of maximum deflection, obtained when both bending and stretching are taken account of, are plotted in fig. 1, where they are compared with the results obtained by the simple theory which ignores stretching.

A solution by Prescott ('Applied Elasticity,' p. 461), assuming that the plate is bent to a portion of a spherical surface, gives a factor 2.21 instead of 1.82 on equation (15). Prescott also considers the case where the radial tension at the edge of the plate is zero and, for  $\sigma=0.25$ , obtains a solution which may be written in the form

$$w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D} \frac{1}{1+0.277 \frac{w_0^2}{h^2}} \quad \dots \quad (20)$$

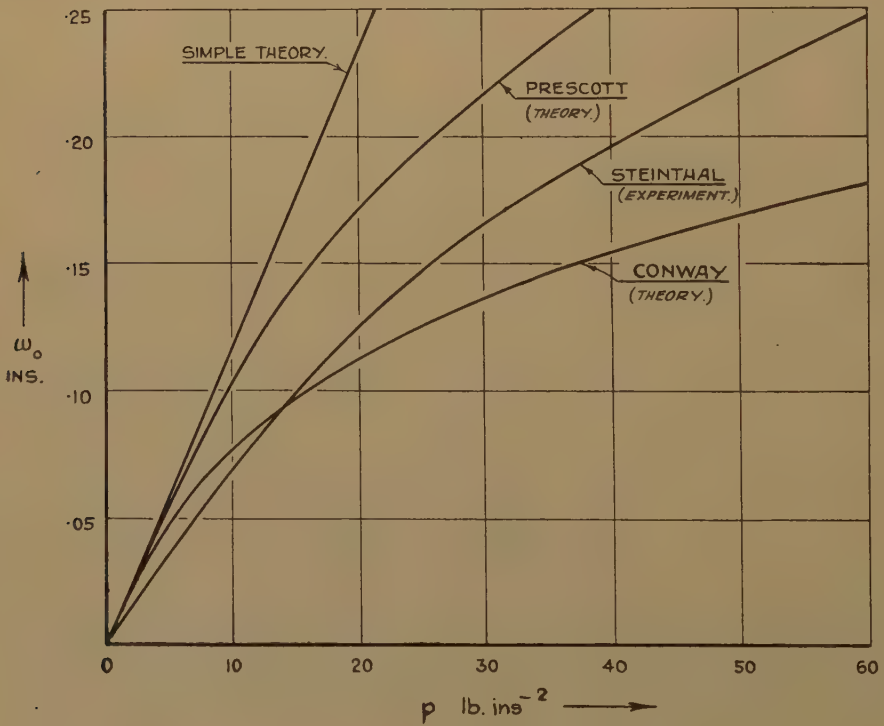
Experimental researches on the deflections of uniformly loaded and simply supported circular plates have been carried out by Steinthal ('Some Experimental Data on the Flexure of Flat Circular Plates within the Electric Limit,' 'Engineering,' vol. xci. p. 677). A typical curve

Fig. 1.



obtained by Steinthal is given in fig. 2, where it is compared with theoretical curves obtained from equations (15), (16), and (20)

Fig. 2.





respectively. All curves relate to a plate of radius 6 in., thickness 0.142 in. and Young's modulus  $27.8 \times 10^6$  lb. per sq. in. Poisson's ratio is assumed to be 0.25. The experimental curve is seen to lie, for the most part, between two of the theoretical curves (equations (15) and (20)), thus indicating that the edges of the plate in the experiments were neither free from radial displacement nor tension.

An approximate solution can also be obtained by a method due to Föppl ('Drang und Zwang,' vol. i p. 345), assuming that the pressure  $p$  is composed of two parts  $p_1$  and  $p_2$ ,  $p_1$  causing pure bending and  $p_2$  pure stretching. For a supported circular plate ignoring stretching, we have for  $\sigma=0.25$

$$w_0 = 0.738 \frac{p_1 a^4}{Eh^3}.$$

From this we have

$$p_1 = 1.355 \frac{Eh^3 w_0}{a^4}. \quad \dots \dots \dots (21)$$

The solution to the problem of a uniformly loaded circular membrane (Hencky, *Zeit. Math. Physik.*, vol. lxiii. p. 311) gives

$$w_0 = 0.662 \sqrt{\frac{p_2 a^4}{Eh}}.$$

From this we have

$$p_2 = 3.447 \frac{Eh w_0^3}{a^4}. \quad \dots \dots \dots (22)$$

Summing equations (21) and (22), we have

$$p = p_1 + p_2 = \frac{w_0 Eh^3}{a^4} \left( 1.355 + 3.447 \frac{w_0^2}{h^2} \right).$$

$$\therefore w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D} \frac{1}{1 + 2.545 \frac{w_0^2}{h^2}}. \quad \dots \dots \dots (23a)$$

For  $\sigma=0.3$ , equation (23a) becomes

$$w_0 = \frac{5+\sigma}{1+\sigma} \frac{pa^4}{64D} \frac{1}{1 + 2.397 \frac{w_0^2}{h^2}}. \quad \dots \dots \dots (23b)$$

The maximum bending moment in the plate occurs at the centre and is given by

$$M_{\max} = -D \left( \frac{\partial^2 w}{\partial r^2} + \frac{\sigma}{r} \frac{\partial w}{\partial r} \right)_{r=0}. \quad \dots \dots \dots (24)$$

From equation (1), this becomes

$$M_{\max} = \frac{3+\sigma}{(5+\sigma)(1-\sigma)} \frac{Eh^3 w_0}{3a^2}. \quad \dots \dots \dots (25)$$

The maximum bending stress is therefore given by

$$f_{\max} = \frac{3+\sigma}{(5+\sigma)(1-\sigma)} \frac{Eh^3w_0}{3a^2} \cdot \frac{6}{h^2} \\ = 2 \frac{3+\sigma}{(5+\sigma)(1-\sigma)} \frac{Ehw_0}{a^2} \quad \dots \quad (26)$$

For  $\sigma=0.25$ , this becomes

$$f_{\max} = 1.650 \frac{Ehw_0}{a^2}, \quad \dots \quad (27a)$$

and, for  $\sigma=0.30$

$$f_{\max} = 1.778 \frac{Ehw_0}{a^2} \quad \dots \quad (27b)$$

The maximum tensile stress occurs at the plate centre and is given by

$$t_{\max} = E \left( \frac{\partial u}{\partial r} \right)_{r=0} \quad \dots \quad (28)$$

Substituting from equation (6), we have for  $\sigma=0.25$

$$t_{\max} = 0.866 \frac{Ew_0^2}{a^2}, \quad \dots \quad (29a)$$

and, for  $\sigma=0.30$

$$t_{\max} = 0.854 \frac{Ew_0^2}{a^2} \quad \dots \quad (29b)$$

The maximum stress in the plate for  $\sigma=0.25$  is therefore given by

$$q_{\max} = f_{\max} + t_{\max} = \frac{Ehw_0}{a^2} \left( 1.650 + 0.866 \frac{w_0}{h} \right), \quad \dots \quad (30a)$$

or, for  $\sigma=0.30$

$$q_{\max} = f_{\max} + t_{\max} = \frac{Ehw_0}{a^2} \left( 1.778 + 0.854 \frac{w_0}{h} \right) \quad \dots \quad (30b)$$

If  $w_0/h$  is neglected in equation (30a) or (30b), the maximum stress is

$$q_{\max} = \frac{3pa^2}{8h^2} (3+\sigma), \quad \dots \quad (31)$$

which is the result obtained from the simple theory neglecting tension. If bending is neglected, we have for  $\sigma=0.25$

$$q_{\max} = 0.866 \frac{Ew_0^2}{a^2} = 0.47 \sqrt[3]{\frac{Ep^2a^2}{h^2}}, \quad \dots \quad (32a)$$

and for  $\sigma=0.30$

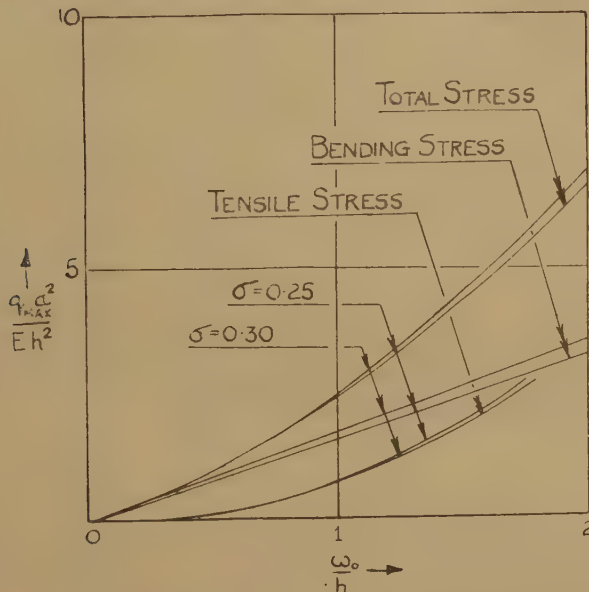
$$q_{\max} = 0.854 \frac{Ew_0^2}{a^2} = 0.44 \sqrt[3]{\frac{Ep^2a^2}{h^2}}, \quad \dots \quad (32b)$$

for the maximum stress in a circular membrane. A more accurate value (Hencky, *Zeit. Math. Physik.* vol. lxiii. p. 311) is

$$q_{\max} = 0.42 \sqrt[3]{\frac{Ep^2a^2}{h^2}} \quad \dots \quad (33)$$

Graphs of maximum stress set up in the circular plate are plotted against maximum deflection in fig. 3, the maximum deflection being plotted in fig. 1.

Fig. 3.



(b) Square plate.

Assume that the displacements  $u, v, w$  at a point  $P(x, y, z)$  may be represented by the following equations. The origin is assumed to be at the plate centre with the co-ordinate axes parallel to the sides.

$$\left. \begin{aligned} u &= k \sin \frac{\pi x}{a} \cos \frac{\pi y}{2a}, \\ v &= k \sin \frac{\pi y}{a} \cos \frac{\pi x}{2a}, \\ w &= w_0 \cos \frac{\pi x}{2a} \cos \frac{\pi y}{2a} \end{aligned} \right\} \dots \dots \dots (34)$$

It will be seen from the first of these two equations that there will be no displacement in the plane of the plate at the centre and at the edges.

It is shown in books on elasticity that the strain energy of bending is given by

$$U_1 = \frac{1}{2} D \int_{-a}^a \int_{-a}^a \left[ \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} \right)^2 - 2(1-\sigma) \left\{ \frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2} - \left( \frac{\partial^2 w}{\partial x \partial y} \right)^2 \right\} \right] \dots \dots \dots (35)$$

Substituting the last of equations (34) in equation (35), we have

$$U_1 = \frac{\pi^4 D}{8} \frac{w_0^2}{a^2}, \dots \dots \dots (36)$$

or, for  $\sigma=0.25$

$$U_1 = 1.083 \frac{Eh^3 w_0^2}{a^2} \quad \dots \quad (37)$$

It can be shown (Timoshenko, 'Theory of Plates and Shells,' p. 345) that the strain energy of stretching  $U_2$  can be represented by

$$\begin{aligned} U_2 = & \frac{6D}{h^2} \int_{-a}^a \int_{-a}^a \left\{ \left( \frac{\partial u}{\partial x} \right)^2 + \frac{\partial u}{\partial x} \left( \frac{\partial w}{\partial x} \right)^2 + \left( \frac{\partial v}{\partial y} \right)^2 + \frac{\partial v}{\partial y} \left( \frac{\partial w}{\partial y} \right)^2 \right. \\ & \left. + \frac{1}{4} \left[ \left( \frac{\partial w}{\partial x} \right)^2 + \left( \frac{\partial w}{\partial y} \right)^2 \right]^2 \right. \\ & \left. + 2\sigma \left[ \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} + \frac{1}{2} \frac{\partial v}{\partial y} \left( \frac{\partial w}{\partial x} \right)^2 + \frac{1}{2} \frac{\partial u}{\partial x} \left( \frac{\partial w}{\partial y} \right)^2 \right] \right. \\ & \left. + \frac{1-\sigma}{2} \left[ \left( \frac{\partial u}{\partial y} \right)^2 + 2 \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} + \left( \frac{\partial v}{\partial x} \right)^2 \right] \right. \\ & \left. + 2 \frac{\partial u}{\partial y} \frac{\partial w}{\partial x} \frac{\partial w}{\partial y} + 2 \frac{\partial v}{\partial x} \frac{\partial w}{\partial x} \frac{\partial w}{\partial y} \right\} dx dy. \quad (38) \end{aligned}$$

Substituting equations (34) in equation (38) and assuming  $\sigma=0.25$ , we have

$$U_2 = Eh \left( 12.700k^2 - 3.729k \frac{w_0^2}{a} + 1.015 \frac{w_0^4}{a^2} \right) \quad \dots \quad (39)$$

The energy of stretching will be a minimum in the equilibrium position, and therefore

$$\frac{\partial U_2}{\partial k} = Eh \left( 25.400k - 3.729 \frac{w_0^2}{a} \right) = 0, \quad \dots \quad (40)$$

from which we obtain

$$k = 0.147 \frac{w_0^2}{a} \quad \dots \quad (41)$$

Substituting in equation (39), we have

$$U_2 = 0.742 \frac{Eh w_0^4}{a^2} \quad \dots \quad (42)$$

Summing equations (37) and (42), the total strain energy of the plate is

$$U = U_1 + U_2 = \frac{Eh^3 w^2}{a^2} \left( 1.083 + 0.742 \frac{w_0^2}{h^2} \right) \quad \dots \quad (43)$$

From the principle of virtual displacements, we have

$$\frac{dU}{dw_0} = p \int_{-a}^a \int_{-a}^a \left( \cos \frac{\pi x}{2a} \cos \frac{\pi y}{2a} \right) dx dy, \quad \dots \quad (44)$$

whence

$$w_0 = 0.748 \frac{p a^4}{E h^3} \frac{1}{1 + 1.37 \frac{w_0^2}{h^2}} \quad \dots \quad (45)$$

For small values of  $w_0/h$ ,

$$w_0 = 0.748 \frac{p a^4}{E h^3}, \quad \dots \quad (46)$$



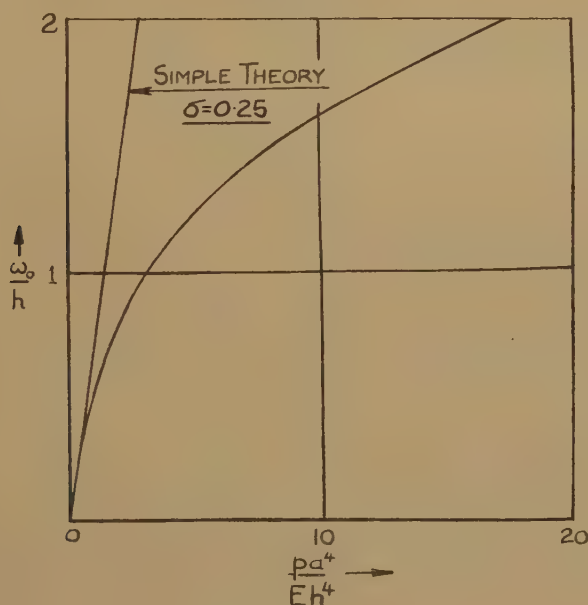
which is the equation obtained when stretching is neglected. A more accurate value for the coefficient in equation (46) is 0.730.

For large values of  $w_0/h$ , we have

$$w_0 = 0.818 \sqrt[3]{\frac{pa^4}{Eh}}, \quad \dots \dots \dots (47)$$

for the maximum deflection of a uniformly loaded square membrane. Föppl ('Drang und Zwang,' vol. i. p. 226) gives the factor in this equation as 0.802. There is, however, an error in his arithmetic and the result, when corrected, is 0.818, as in equation (47). A graph of maximum deflection of a uniformly loaded square plate, obtained when both bending and stretching are taken account of, is plotted in fig. 4. A comparison is made with the results obtained by the simple theory which ignores stretching.

Fig. 4.



Prescott ('Applied Elasticity,' p. 478) has given a solution to the problem using a different method, his result in the nomenclature of this paper being

$$w_0 = 0.749 \frac{pa^4}{Eh^3} \frac{1}{1 + 1.289 \frac{w_0^2}{h^2}}, \quad \dots \dots \dots (48)$$

which is in fair agreement with equation (45). Föppl's method may also be used, and, by the use of this, Timoshenko ('Theory of Plates and

Shells,' p. 350) has given a solution which, when corrected for the error mentioned above, becomes

$$w_0 = 0.730 \frac{pa^4}{Eh^3} \frac{1}{1 + 1.336 \frac{w_0^2}{h^2}}, \quad \dots \quad (49)$$

which agrees quite well with equation (45).

The maximum bending moment in the plate occurs at the centre and is given by

$$M_{\max} = -D \left( \frac{\partial^2 w}{\partial x^2} + \sigma \frac{\partial^2 w}{\partial y^2} \right)_{x=0, y=0} \quad \dots \quad (50)$$

From equation (34), this becomes

$$\begin{aligned} M_{\max} &= \frac{\pi^2 D (1 + \sigma) w_0}{4a^2} \\ &= \frac{\pi^2 E h^3 w_0}{48(1 - \sigma)a^2}. \quad \dots \quad (51) \end{aligned}$$

The maximum bending stress is therefore given by

$$\begin{aligned} f_{\max} &= \frac{\pi^2 E h^3 w_0}{48(1 - \sigma)a^2} \cdot \frac{6}{h^2} \\ &= \frac{\pi^2 E h w_0}{8(1 - \sigma)a^2}. \quad \dots \quad (52) \end{aligned}$$

For  $\sigma = 0.25$ , this becomes

$$f_{\max} = 1.645 \frac{E h w_0}{a^2}. \quad \dots \quad (53)$$

The maximum tensile stress occurs at the plate centre and is given by

$$\begin{aligned} t_{\max} &= \frac{E}{1 - \sigma} \left( \frac{\partial u}{\partial x} \right)_{x=0} \\ &= 0.616 \frac{E w_0^2}{a^2}. \quad \dots \quad (54) \end{aligned}$$

The maximum resultant stress in the plate is therefore given by

$$q_{\max} = f_{\max} + t_{\max} = \frac{E h w_0}{a^2} \left( 1.645 + 0.616 \frac{w_0}{h} \right). \quad \dots \quad (55)$$

If  $w_0/h$  is neglected in equation (55), the maximum stress is

$$\begin{aligned} q_{\max} &= 1.645 \frac{E h w_0}{a^2} \\ &= 1.230 \frac{pa^2}{h^2}, \quad \dots \quad (56) \end{aligned}$$

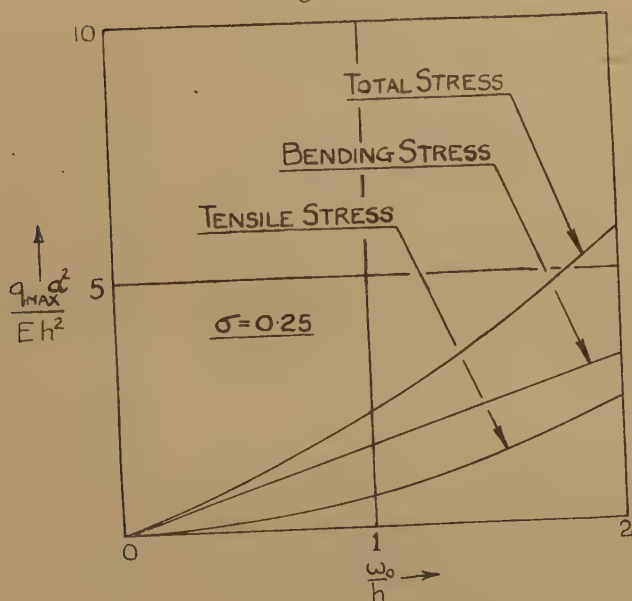
from equation (46).

For large values of  $w_0/h$ , the maximum stress is

$$q_{\max} = 0.616 \frac{E w_0^2}{a^2}$$

$$= 0.412 \sqrt[3]{\frac{p^2 E a^2}{h^2}} \dots \dots \dots (57)$$

Fig. 5.



Graphs of maximum stress set up in the square plate are plotted against maximum deflection in fig. 5, the maximum deflection being plotted in fig. 4.

# LXXIX. The Large Deflections of Rectangular Membranes and Plates.

By H. D. CONWAY, M.A., Ph.D.\*

[Received November 2, 1945.]

## 1. Introduction.

IN an earlier paper <sup>(1)</sup>, calculations for the maximum deflection and stress in uniformly loaded square membranes and plates were given, using an approximate method due to A. and L. Föppl <sup>(2)</sup>. It was decided to extend the calculations to the more general case of the rectangle, and the results given in this paper are applicable to rectangular membranes and plates

\* Communicated by the Author.

having length-breadth ratios of from 1-4, this being considered to cover most practical cases. The edges of the plates are assumed to be fixed in position but allowed to rotate on loading—a condition which is frequently met with in practice. Finally, the results given by the method are compared with those given by an approximate theory due to J. Prescott <sup>(3)</sup>.

## 2. Analysis.

Fig. 1.

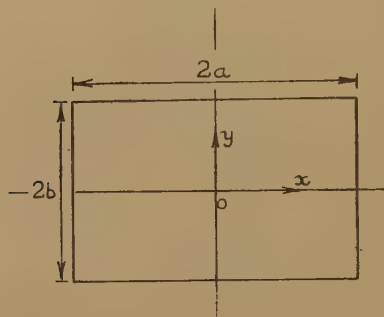


Fig. 1 shows a diagram of a membrane or plate of sides  $2a$  and  $2b$ , the origin of co-ordinates being taken at the centre.

### (a) The Uniformly Loaded Membrane.

The strain energy  $U$  of a rectangular membrane of sides  $2a$  and  $2b$  due solely to stretching shown by Föppl to be

$$U = \frac{Eh}{2(1-\sigma^2)} \int_{-a}^a \int_{-b}^b [\epsilon_x^2 + \epsilon_y^2 + 2\sigma\epsilon_x\epsilon_y + \frac{1}{2}(1-\sigma)\gamma_{xy}^2] dx dy, \quad (1)$$

where

$\epsilon_x$  = strain in direction of  $x$ ,

$\epsilon_y$  = " " " " "  $y$ ,

$\gamma_{xy}$  = shearing strain,

$h$  = thickness of membrane,

$\sigma$  = Poisson's ratio,

$E$  = Young's modulus.

If the displacements at a point  $P(x, y, z)$  in the membrane are  $u, v$  and  $w$  in the directions of the  $x, y$  and  $z$  axes respectively, we may write

$$\left. \begin{aligned} \epsilon_x &= \frac{\partial u}{\partial x} + \frac{1}{2} \left( \frac{\partial w}{\partial x} \right)^2, \\ \epsilon_y &= \frac{\partial v}{\partial y} + \frac{1}{2} \left( \frac{\partial w}{\partial y} \right)^2, \\ \gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} + \frac{\partial w}{\partial x} \cdot \frac{\partial w}{\partial y} \end{aligned} \right\} \dots \dots \dots (2)$$



The following expressions are assumed for the displacement  $u, v, w$ , it being observed that they vanish at the boundary of the membrane:—

$$\left. \begin{aligned} u &= c_1 \sin \frac{\pi x}{a} \cos \frac{\pi y}{2b}, \\ v &= c_2 \sin \frac{\pi y}{b} \cos \frac{\pi x}{2a}, \\ w &= w_0 \cos \frac{\pi x}{2a} \cos \frac{\pi y}{2b} \end{aligned} \right\} \dots \dots \dots (3)$$

Substituting equations (3) in equations (2), we have

$$\left. \begin{aligned} \epsilon_x &= \frac{\pi^2}{8a^2} w_0^2 \sin^2 \frac{\pi x}{2a} \cos^2 \frac{\pi y}{2b} + \frac{\pi}{a} c_1 \cos \frac{\pi x}{a} \cos \frac{\pi y}{2b}, \\ \epsilon_y &= \frac{\pi^2}{8b^2} w_0^2 \cos^2 \frac{\pi x}{2a} \sin^2 \frac{\pi y}{2b} + \frac{\pi}{b} c_2 \cos \frac{\pi x}{2a} \cos \frac{\pi y}{b}, \\ \gamma_{xy} &= \frac{\pi^2}{16ab} w_0^2 \sin \frac{\pi x}{a} \sin \frac{\pi y}{b} - \frac{\pi}{2b} c_1 \sin \frac{\pi x}{a} \sin \frac{\pi y}{2b} - \frac{\pi}{2a} c_2 \sin \frac{\pi x}{2a} \sin \frac{\pi y}{b} \end{aligned} \right\} \dots \dots \dots (4)$$

Therefore, we have

$$\left. \begin{aligned} \int_{-a}^a \int_{-b}^b \epsilon_x^2 dx dy &= \frac{9\pi^4}{1024} \frac{b}{a^3} w_0^4 - \frac{\pi^2}{3} \frac{b}{a^2} w_0^2 c_1 + \pi^2 \frac{b}{a} c_1^2, \\ \int_{-a}^a \int_{-b}^b \epsilon_y^2 dx dy &= \frac{\pi^4}{1024ab} w_0^4 + \frac{\pi^2}{12b} w_0^2 c_1 + \frac{\pi^2}{12a} w_0^2 c_2 + \frac{16}{9} c_1 c_2, \\ \int_{-a}^a \int_{-b}^b \gamma_{xy}^2 dx dy &= \frac{\pi^4}{256ab} w_0^4 - \frac{\pi^2}{6b} w_0^2 c_1 - \frac{\pi^2}{6a} w_0^2 c_2 + \frac{\pi^2 a}{4b} c_1^2 \\ &\quad + \frac{\pi^2 b}{4a} c_2^2 + \frac{32}{9} c_1 c_2 \end{aligned} \right\} \dots \dots \dots (5)$$

The expression for the strain energy thus becomes

$$\begin{aligned} U &= \frac{Eh}{2(1-\sigma^2)} \left\{ \frac{\pi^4}{1024ab} w_0^4 \left( 9 \frac{a^2}{b^2} + 9 \frac{b^2}{a^2} + 2 \right) - \frac{\pi^2}{6} w_0^2 c_1 \left( 2 \frac{b}{a^2} + \frac{1-3\sigma}{2b} \right) \right. \\ &\quad - \frac{\pi^2}{6} w_0^2 c_2 \left( 2 \frac{a}{b^2} + \frac{1-3\sigma}{2a} \right) + \pi^2 c_1^2 \left( \frac{b}{a} + \frac{1-\sigma}{8} \frac{a}{b} \right) \\ &\quad \left. + \pi^2 c_2^2 \left( \frac{a}{b} + \frac{1-\sigma}{8} \frac{b}{a} \right) + \frac{16(1+\sigma)}{9} c_1 c_2 \right\}, \dots \dots \dots (6) \end{aligned}$$

or, for  $\sigma=0.25$ ,

$$\begin{aligned} U &= \frac{8Eh}{15} \left\{ \frac{\pi^4}{1024ab} w_0^4 \left( 9 \frac{a^2}{b^2} + 9 \frac{b^2}{a^2} + 2 \right) - \frac{\pi^2}{6} w_0^2 c_1 \left( 2 \frac{b}{a^2} + \frac{1}{8b} \right) \right. \\ &\quad - \frac{\pi^2}{6} w_0^2 c_2 \left( 2 \frac{a}{b^2} + \frac{1}{8a} \right) + \pi^2 c_1^2 \left( \frac{b}{a} + \frac{3a}{32b} \right) \\ &\quad \left. + \pi^2 c_2^2 \left( \frac{a}{b} + \frac{3b}{32a} \right) + \frac{20}{9} c_1 c_2 \right\}, \dots \dots \dots (7) \end{aligned}$$



applying equation (11), the maximum deflection  $w_0$  can be obtained in the form

$$w_0 = \gamma \sqrt[3]{\frac{pa^4}{Eh}}, \quad \dots \dots \dots (12)$$

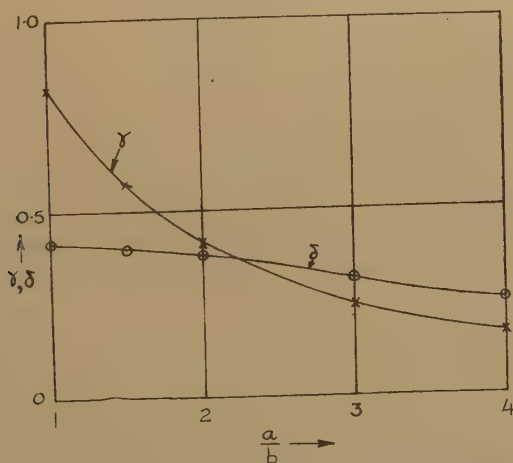
where  $\gamma$  is a constant depending on the ratio  $a/b$ . This constant has been calculated for  $a/b = 1.0, 1.5, 2.0, 3.0$  and  $4.0$ , and is given below in Table II.

TABLE II.

$w_0 = \gamma \sqrt[3]{\frac{pa^4}{Eh}}$	$\frac{a}{b} =$	1.0	1.5	2.0	3.0	4.0
	$\gamma =$	0.818	0.569	0.411 <sub>5</sub>	0.248 <sub>5</sub>	0.171

A graph of  $\gamma$  against  $a/b$  is plotted in fig. 2 from which the value of the constant  $\gamma$  in equation (12) can be obtained for any value of  $a/b$ .

Fig. 2.



The tensile strains at the centre of the membrane are given by

$$\left. \begin{aligned} \epsilon_x &= \left( \frac{\partial u}{\partial x} \right)_{x=y=0} = \frac{\pi}{a} c_1, \\ \epsilon_y &= \left( \frac{\partial v}{\partial y} \right)_{x=y=0} = \frac{\pi}{b} c_2 \end{aligned} \right\} \dots \dots \dots (13)$$

The tensile stresses are given by

$$\left. \begin{aligned} f_x &= \frac{\pi E}{1-\sigma^2} \left( \frac{c_1}{a} + \sigma \frac{c_2}{b} \right), \\ f_y &= \frac{\pi E}{1-\sigma^2} \left( \frac{c_2}{b} + \sigma \frac{c_1}{a} \right) \end{aligned} \right\} \dots \dots \dots (14)$$

or, for  $\sigma=0.25$ ,

$$\left. \begin{aligned} f_x &= \frac{16\pi E}{15} \left( \frac{c_1}{a} + \frac{c_2}{4b} \right), \\ f_y &= \frac{16\pi E}{15} \left( \frac{c_2}{b} + \frac{c_1}{4a} \right) \end{aligned} \right\} \dots \dots \dots (15)$$

For  $a/b > 1$ , the greatest stress is  $f_y$ , and using the values of  $c_1$  and  $c_2$  given in Table I. and the values of  $w_0$  given in Table II., the maximum stress may be written in the form

$$f_y = \delta \sqrt[3]{\frac{p^2 E a^2}{h^2}}, \dots \dots \dots (16)$$

where  $\delta$  is a constant depending on the ratio  $a/b$ . Values of  $\delta$  have been calculated for  $a/b=1.0, 1.5, 2.0, 3.0$  and  $4.0$ , and are given below in Table III.

TABLE III.

$f_y = \delta \sqrt[3]{\frac{p^2 E a^2}{h^2}}$	$\frac{a}{b} =$	1.0	1.5	2.0	3.0	4.0
$\delta =$		0.412	0.394 <sub>5</sub>	0.383	0.310 <sub>5</sub>	0.260

A graph of  $\delta$  against  $a/b$  is given in fig. 2, from which the value of the constant  $\delta$  in equation (16) can be found for any intermediate value of  $a/b$ .

(b) *The Uniformly Loaded Plate.*

Having calculated the maximum deflection and stress in rectangular membranes, it was decided to use the results in the calculation of the approximate maximum deflection and stress in uniformly loaded thin rectangular plates secured in position but not direction at the edges. This was done by the use of a device due to Föppl.

Consider the uniform pressure  $p$  on the plate as consisting of two separate pressures  $p_1$  and  $p_2$ , the former of which causes bending and the latter pure stretching of the plate. The pressures  $p_1$  and  $p_2$  are then calculated in terms of the maximum deflection  $w_0$  and are added to give expressions for uniformly loaded thin plates in which both bending and stretching takes place. Expressions for the maximum stress are deduced in a similar manner and the method will be illustrated by an example. Before proceeding, however, it is necessary to obtain expressions for the maximum deflection and stress due to bending in uniformly loaded and simply supported rectangular plates.



S. Timoshenko <sup>(4)</sup> has given the maximum deflection and stress due to bending in the following forms :—

$$\left. \begin{aligned} w_0 &= \frac{5}{2}(1-\sigma^2) \frac{p_1 a^4}{Eh^3} - \frac{768(1-\sigma^2) p_1 a^4}{\pi^5 Eh^3} \sum_{m=1,3,5,\dots}^{\infty} \frac{(-1)^{\frac{m-1}{2}}}{m^5} \cdot \frac{\alpha_m \tanh \alpha_m + 2}{2 \cosh \alpha_m} \\ f_v &= \frac{3p_1 a^2}{h^2} + 24\pi^2(1-\sigma^2) \frac{p_1 a^2}{h^2} \sum_{m=1,3,5,\dots}^{\infty} (-1)^{\frac{m-1}{2}} m^2 \left( A_m - \frac{2\sigma}{1-\sigma} B_m \right) \end{aligned} \right\} \quad (17)$$

$$\text{where } \alpha_m = m \cdot \frac{2a}{\pi b}, \quad A_m = \frac{-2(\alpha_m \tanh \alpha_m + 2)}{\pi^5 m^5 \cosh \alpha_m}, \quad B_m = \frac{2}{\pi^5 m^5 \cosh \alpha_m}.$$

Unfortunately, Timoshenko's tabulated values of  $w_0$  and  $f_v$  for various values of  $a/b$  are calculated for Poisson's ratio  $\sigma=0.3$ , and they had to be re-calculated for  $\sigma=0.25$ . However, the series are very rapidly convergent and the labour involved in the calculation was quite small. The maximum deflection and stress due to bending may be expressed in the forms

$$\left. \begin{aligned} w_0 &= \zeta \frac{p_1 a^4}{Eh^3}, \\ f_v &= \xi \frac{p_1 a^2}{h^2} \end{aligned} \right\} \quad \dots \dots \dots (18)$$

where  $\zeta$  and  $\xi$  are constants dependent upon the ratio  $a/b$ . Values of  $\zeta$  and  $\xi$  were calculated for  $a/b=1, 1.5, 2, 3$  and  $4$ , and are given below in Table IV.

TABLE IV.

$w_0 = \zeta \frac{p_1 a^4}{Eh^3}$	$\frac{a}{b} =$	1.0	1.5	2.0	3.0	4.0
$\zeta =$		0.730	0.274	0.114	0.027	0.009
$f_v = \xi \frac{p_1 a^2}{h^2}$						
$\xi =$		1.099	0.850	0.605	0.315	0.185

Föppl's method will be illustrated for the ratio  $a/b=1.5$ .

From Table IV., we have for the bending deflection of a plate having  $a/b=1.5$

$$\begin{aligned} w_0 &= 0.274 \frac{p_1 a^4}{Eh^3}, \\ \therefore p_1 &= 3.64 \frac{w_0 Eh^3}{a^4}. \end{aligned} \quad \dots \dots \dots (19)$$

From Table II., we have for the deflection of a membrane having  $a/b=1.5$

$$w_0 = 0.569 \sqrt[3]{\frac{p_2 a^4}{Eh}}$$

$$\therefore p_2 = 5.42 \frac{w_0^3 E h}{a^4} \quad \dots \dots \dots (20)$$

Adding equations (19) and (20), we have

$$p = p_1 + p_2 = \frac{w_0 E h^3}{a^4} \left( 3.64 + 5.42 \frac{w_0^2}{h^2} \right),$$

$$\therefore \frac{p a^4}{E h^4} = \frac{w_0}{h} \left( 3.64 + 5.42 \frac{w_0^2}{h^2} \right) \quad \dots \dots \dots (21)$$

From this cubic equation, the maximum deflection  $w_0$  may be calculated.

From Table IV., the maximum bending stress in a thin plate having  $a/b=1.5$  is

$$\text{Maximum bending stress} = 0.850 \frac{p_1 a^2}{h^2}$$

$$= 3.10 \frac{w_0 E h}{a^2} \quad \dots \dots \dots (22)$$

from equation (19).

From Table III., the maximum tensile stress in a membrane having  $a/b=1.5$  is

$$\text{Maximum tensile stress} = 0.394_5 \sqrt[3]{\frac{p_2^2 E a^2}{h^2}}$$

$$= 1.22 \frac{w_0^2 E}{a^2} \quad \dots \dots \dots (23)$$

from equation (20).

The two stresses in equations (22) and (23) both occur at the centre of the plate and are in the same direction (parallel to the short side). Adding the stresses, we have

$$\text{Total stress } f_y = \frac{w_0 E h}{a^2} \left( 3.10 + 1.22 \frac{w_0}{h} \right)$$

$$\text{or } \frac{f_y a^2}{E h^2} = \frac{w_0}{h} \left( 3.10 + 1.22 \frac{w_0}{h} \right) \quad \dots \dots \dots (24)$$

From this quadratic equation, the maximum stress  $f_y$  can be calculated, the maximum deflection  $w_0$  having first been calculated from equation (21). In general, for any value of  $a/b$ , we may write,

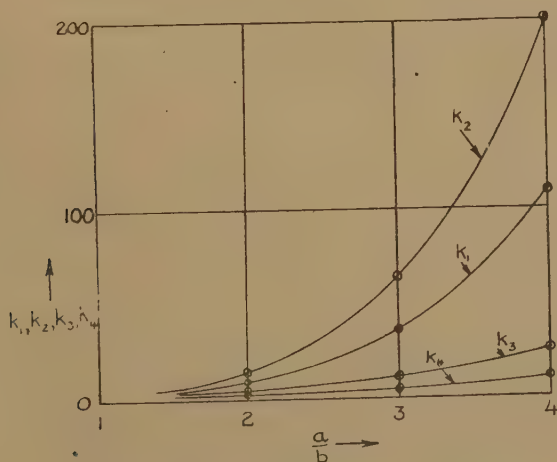
$$\left. \begin{aligned} \frac{p a^4}{E h^4} &= \frac{w_0}{h} \left( k_1 + k_2 \frac{w_0^2}{h^2} \right), \\ \frac{f_y a^2}{E h^2} &= \frac{w_0}{h} \left( k_3 + k_4 \frac{w_0}{h} \right) \end{aligned} \right\} \quad \dots \dots \dots (25)$$

where  $k_1, k_2, k_3$  and  $k_4$  are constants dependent upon  $a/b$ . It will be noted that if  $\frac{w_0}{h}$  is small, the tension terms are negligible and the ordinary theory results, which neglects stretching of the middle surface of the plate. Conversely, if  $\frac{w_0}{h}$  is large, the bending terms are negligible and the plate may be treated as a membrane. Values of  $k_1, k_2, k_3$  and  $k_4$  have been calculated for  $a/b=1.0, 1.5, 2.0, 3.0$  and  $4.0$ , and are given below in Table V.

TABLE V.

$\frac{pa^4}{Eh^4}$	$\frac{a}{b}=$	1.0	1.5	2.0	3.0	4.0
$= \frac{w_0}{h} \left( k_1 + k_2 \frac{w_0^2}{h^2} \right)$	$k_1=$	1.37	3.64	8.78	36.81	110.96
	$k_2=$	1.83	5.42	14.35	65.22	201.02
$\frac{f_y a^2}{Eh^2}$	$k_3=$	1.51	3.10	5.31	11.59	20.54
	$k_4=$	0.62	1.22	2.26	5.03	8.93

Fig. 3.



Graphs of  $k_1, k_2, k_3$  and  $k_4$  are plotted against  $a/b$  in fig. 3 and graphs of  $\frac{pa^4}{Eh^4}$  and  $\frac{f_y a^2}{Eh^2}$  against  $\frac{w_0}{h}$  for  $a/b=1.0, 1.5, 2.0, 3.0$  and  $4.0$  in figs. 4 and 5

respectively. The straight lines in figs. 4 and 5 are given by the simplified theory which neglects stretching of the middle surface.

Fig. 4.

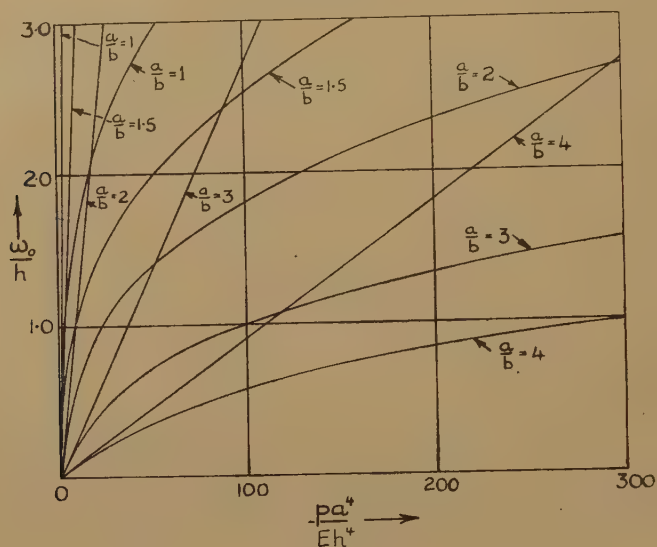
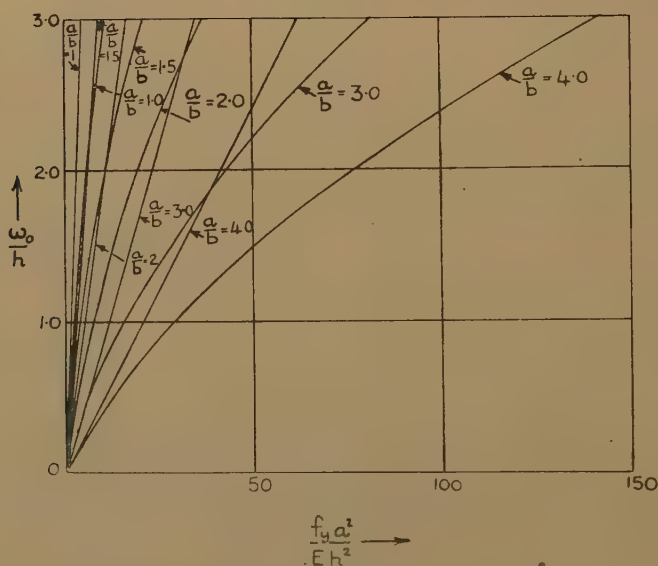


Fig. 5.



The maximum deflections and stresses in uniformly loaded rectangular plates, having edges fixed in position but not in direction, have been obtained by J. Prescott, using a different method to that given in this



paper. His solutions, using the notation of this paper, are given by

$$\left. \begin{aligned} p &= \frac{\pi^6 E h^3 w_0}{1024 (1-\sigma^2)} \left\{ \frac{1}{3} \left( \frac{1}{a^2} + \frac{1}{b^2} \right)^2 + \frac{w_0^2}{h^2} \left[ \frac{\sigma}{a^2 b^2} + \frac{3-\sigma^2}{4} \left( \frac{1}{a^4} + \frac{1}{b^4} \right) \right] \right\} \\ f_y &= \frac{\pi^2 E w_0}{32 (1-\sigma^2)} \left\{ \frac{(2-\sigma^2) w_0 + 4h}{a^2} + \frac{\sigma(w_0 + 4h)}{b^2} \right\} \end{aligned} \right\} \quad (26)$$

Assuming  $\sigma=0.25$ , equations (26) may be put in the form of equations (25) for various values of  $a/b$ . Calculations of the constants  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  from Prescott's solution have been made and are compared in Table VI. with the results previously given in Table V.

TABLE VI.

$\frac{a}{b}$	1.0	1.5	2.0	3.0	4.0
$k_1$ (Prescott) ..	1.34	3.53	8.34 <sub>5</sub>	33.38	96.47
$k_1$ (Föppl) ..	1.37	3.64	8.78	36.81	110.96
$k_2$ (Prescott) ..	1.72	5.02	13.50	62.56	193.01
$k_2$ (Föppl) ..	1.83	5.42	14.35	65.22	201.02
$k_3$ (Prescott) ..	1.64	3.29	5.59	12.17	21.38
$k_3$ (Föppl) ..	1.51	3.10	5.31	11.59	20.54
$k_4$ (Prescott) ..	0.72	1.52	2.63	5.82	10.53
$k_4$ (Föppl) ..	0.62	1.22	2.26	5.03	8.93

It is seen that there is quite good agreement between the results obtained from Prescott's and Föppl's analyses, Prescott's values for the deflection and stress constants being respectively lower and higher than Föppl's.

### 3. Summary.

1. Expressions for the deflection  $w_0$  and maximum stress  $f_y$  in uniformly loaded rectangular membranes are given :

$$w_0 = \gamma \sqrt[3]{\frac{p a^4}{E h}},$$

$$f_y = \delta \sqrt[3]{\frac{p^2 E a^2}{h^2}},$$

where  $p$  is the pressure, 2 is the longer side of the membrane,  $h$  is the thickness and  $E$  is Young's modulus. The constants  $\gamma$  and  $\delta$  are plotted against the ratio of the sides of the plate  $\frac{a}{b}$  ( $\geq 1$ ) in fig. 2.

2. Expressions are given for the maximum deflection  $w_0$  and maximum stress  $f_y$  in uniformly loaded rectangular plates secured in position but not in direction at the edges :

$$\frac{pa^4}{Eh^4} = \frac{w_0}{h} \left( k_1 + k_2 \frac{w_0^2}{h^2} \right),$$

$$\frac{f_y a^2}{Eh^2} = \frac{w_0}{h} \left( k_3 + k_4 \frac{w_0}{h} \right).$$

The constants  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  are plotted against  $a/b$  in fig. 3, and graphs of  $\frac{pa^4}{Eh^4}$  and  $\frac{f_y a^2}{Eh^2}$  against  $\frac{w_0}{h}$  for  $\frac{a}{b} = 1.0, 1.5, 2.0, 3.0$  and  $4.0$  in figs. 4 and 5 respectively.

3. The results of the analysis are compared with those from a theory due to Prescott, quite good agreement being found.

#### *References.*

- (1) H. D. Conway, *Phil. Mag.* (7) vol. xxxvii, p. 756 (1946).
- (2) A. and L. Föppl, 'Drang und Zwang,' Oldenburg, Munich, vol. i. p. 226 (1920).
- (3) J. Prescott, 'Applied Elasticity,' Longmans, Green and Co., pp. 474-478 (1920).
- (4) S. Timoshenko, 'Theory of Plates and Shells,' McGraw-Hill Book Company Inc., pp. 125-133 (1940).

#### LXXX. *A Contribution to Linear Network Analysis.*

By D. K. C. MACDONALD, M.A., A.Inst.P., Captain, R.E.M.E.,  
Military College of Science\*.

[Received November 12, 1945.]

#### 1. *Introduction.*

A collection of inter-connected fundamental physical elements (*i. e.* electrical or mechanical) may be called a network. The individual elements will be characterized by their response to an applied stimulus. If the response of such an element is always linearly proportional to the stimulus or to its derivatives or integrals, then we call such an element a linear element. That is, denoting the stimulus by  $e(t)$  and the response by  $i(t)$  (*e. g.*, voltage and current in an electrical element), such elements are characterized by the equation  $i(t) = AD^n e(t) \dagger$  ( $n$  any integer, positive

\* Communicated by the Author.

†  $D \equiv \frac{d}{dt}$  (an operator).  $A$  is a constant of proportionality.

or negative, including zero). An aggregate of such elements is known as a "linear network."

In order to specify the performance of such a network as a whole it is desirable to know its response to some standard type of stimulus from which response, generally unique for a given network, it should be possible to derive the theoretical response to any type of stimulus. The concept of the "frequency response" of a network to an applied sinusoidal type of stimulus is widely known and of very frequent application in electrical engineering and physical science. The response of a network to an applied "step" or "Heaviside" stimulus is hardly less familiar now and will be outlined below. A third method of specifying the network rests in the application of a pulse stimulus; this approach has also received attention in the past and will be mentioned in what follows.

This essay first presents a short unifying summary of these methods with emphasis of certain useful results, and then continues to suggest the possible employment of a response specification and investigation through the introduction of an applied stimulus of exponential form\*, and the initial groundwork of the analysis, etc., based upon the use of such a parameter.

## 2. "Frequency Response" or "Complex Admittance."

A network may be specified in terms of its response  $i(t)$  to an applied stimulus of sinusoidal (or "cisoidal" form)† *assumed applied for all time  $t$* . Thus if

$$i(t) = Y(\omega) \cdot e^{j\omega t}$$

(where  $j = \sqrt{-1}$ ;  $\omega = 2\pi f$ , and  $f$  is the frequency;  $e(t) = e^{j\omega t}$ , the applied "cisoidal" stimulus). Then the function  $Y(\omega)$  so defined is known as the "frequency response" or "complex admittance" of the network. It is, for example, well known that for an electrical circuit composed of resistance  $R$  in series with inductance  $L$  that:

$$i(t) = \left( \frac{1}{R + j\omega L} \right) \cdot e^{j\omega t},$$

i. e.

$$Y(\omega) = \left( \frac{1}{R + j\omega L} \right),$$

is the "complex admittance."

\* To avoid possible confusion with the proposal for a form of pulse stimulus suggested and discussed by F. F. Roberts and J. C. Simmonds<sup>(1)</sup> and called "recurrent exponential (and probability function) pulse(s)," this will be referred to as an "Exponential Signal."

† The word cisoidal has been coined (*e. g.* Bloch<sup>(2)</sup>) to denote the use of the complex function  $e^{j\omega t} (\equiv \cos \omega t + j \sin \omega t)$  to represent the application of a purely real, physical sinusoidal stimulus; simplification of the analysis frequently results.

The analysis of networks in terms of frequency response has been discussed in the past in great detail (*e.g.* (5), (6)). From a complete knowledge of  $Y(\omega)$  (*i.e.* for all values of  $\omega$ ) it is possible to express the response of the network to *any* applied stimulus through the medium of the Fourier transform (or for a stimulus applied at  $t=0$ , say, through the Laplace transform, which in some respects is preferable). It seems opportune here to emphasize that the *application* of the Fourier transform is *not* confined to problems involving inherent periodicity either in the network itself or in the applied stimulus. From the practical point of view  $Y(\omega)$  can be determined for an arbitrary network, in principle, by applying a sinusoidal input from a signal generator, and observing the relationship of the response to the input when conditions have become steady.

If, then, a *general* input, say  $e(t)$ , be applied it can be shown (*e.g.* (3), (4), (5), (6), (7)) that the response  $i(t)$  is

$$\begin{aligned} i(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) \cdot E(\omega) \cdot e^{j\omega t} d\omega, \\ \text{where } E(\omega) &= \int_{-\infty}^{\infty} e(t) \cdot e^{-j\omega t} dt, \\ \text{or } i(t) &= \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} Y(p) \cdot E(p) e^{pt} dp^*, \\ \text{where } E(p) &= \int_0^{\infty} e(t) \cdot e^{-pt} dt \quad (\text{if } e(t) \equiv 0; t < 0 \text{ and } i(t) = 0; t = 0+). \end{aligned} \quad (1)$$

Thus, for example, if

$$e(t) = \frac{2}{4+t^2} \dagger, \quad (2)$$

then 
$$E(\omega) = \int_{-\infty}^{\infty} \frac{2}{4+t^2} \cdot e^{-j\omega t} dt = \pi e^{-2|\omega|}.$$

If, then, 
$$Y(\omega) = \frac{\sin \alpha \omega}{\alpha \omega} \cdot e^{-j\alpha \omega} \ddagger,$$

\*  $Y(p)$  is the function characterizing the network if we replace the operator  $\frac{d}{dt}$  by  $p$ , wherever it arises.

† This function arises as the rate of change of azimuth of a particle moving in a straight line with constant speed.

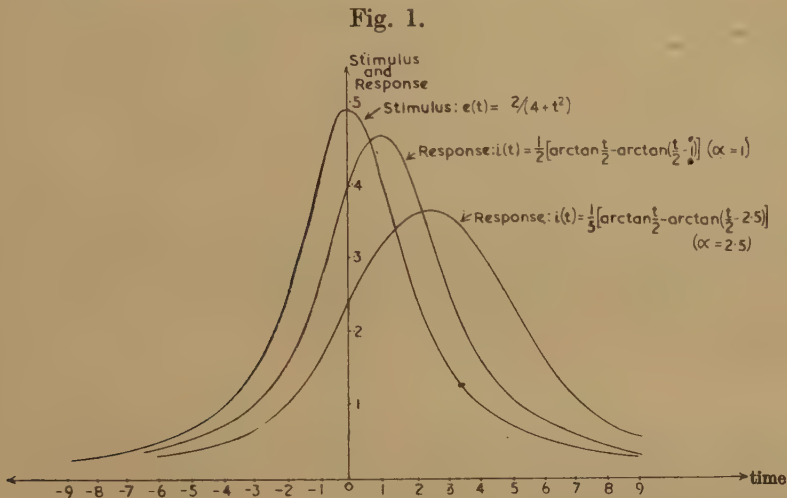
‡ This network function, belonging to a family of functions discussed by the writer<sup>(8)</sup> in another paper, is "physically reliable" in the sense that the choice of amplitude response  $\frac{\sin \alpha \omega}{\alpha \omega}$  stipulated is compatible with linear phase response  $e^{-j\alpha \omega}$  stipulated. That is, the response does not outrage the physical intuition that the response cannot precede in time the application of the stimulus.



$$i(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \pi \cdot e^{-2|\omega|} \cdot \frac{\sin \alpha \omega}{\alpha \omega} \cdot e^{-j\alpha \omega} \cdot e^{j\omega t} \cdot d\omega,$$

$$= \frac{1}{2\alpha} \left\{ \arctan \frac{t}{2} - \arctan \left( \frac{t}{2} - \alpha \right) \right\} \quad \dots \quad (3)$$

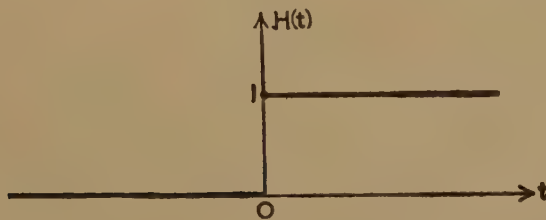
$i(t)$  and  $e(t)$  are sketched together in fig. 1 for two values of  $\alpha$ .



### 3. "Indicial Admittance."

If a unit "step" or "Heaviside" input  $H(t)$  (see fig. 2) be applied to

Fig. 2.



the network, then the response generally denoted by  $A(t)$  is known as the "indicial admittance" of the network (*e. g.* (9)). From a knowledge of the indicial admittance the response can be derived for any linear network *ab initio* to any applied stimulus  $e(t)$  at time  $t=0$  by "building up" or synthesizing the required stimulus from a succession of step-functions. This is expressible in the Duhamel integral (*e. g.* (10)):

$$i(t) = A(t) \cdot e(0) + \int_0^t A(t-\lambda) \cdot \frac{\partial e(\lambda)}{\partial \lambda} \cdot d\lambda \quad \dots \quad (4)$$

or 
$$i(t) = A(0) \cdot e(t) + \int_0^t \frac{\partial}{\partial t} (A(t-\lambda)) \cdot e(\lambda) \cdot d\lambda. \quad . \quad . \quad . \quad (5)$$

—if in addition we exclude the possibility of  $A(0) \neq 0$ , then a convenient form is

$$i(t) = \int_0^t A'(t-\lambda) \cdot e(\lambda) \cdot d\lambda. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

We may also relate the “indicial admittance” to the complex admittance by considering, for example, the response to a step-function through the Fourier transform. In general, the complex Fourier transform (or Laplace transform) is employed for such problems, since, strictly speaking, the integral  $E(\omega)$  does not converge when

$$\begin{aligned} e(t) &= 1; \quad t > 0, \\ &= 0; \quad t < 0. \end{aligned}$$

It may, however, be shown in a variety of ways that if  $E(\omega)$  be defined as

$$E(\omega) = \frac{1}{j\omega} + \pi \cdot \delta(\omega) \quad . \quad . \quad . \quad . \quad . \quad (7)$$

(where  $\delta(\omega)$  is the “impulse” Dirac function such that  $\int_{-\infty}^{\infty} \delta(\omega) \cdot d\omega = 1$  and  $\delta(\omega) \equiv 0$ ;  $\omega \neq 0$ ), then this will yield  $i(t)$  when employed in the Fourier transform for an applied step-function.

That is to say, we may write formally

$$\int_0^{\infty} e^{-j\omega t} dt \triangleq \frac{1}{j\omega} + \pi \delta(\omega) \quad . \quad . \quad . \quad . \quad . \quad (7a)$$

(using the symbol  $\triangleq$  to mean “is defined as equal to”).

Similarly we have

$$\int_{-\infty}^{\infty} e^{-j\omega t} dt \triangleq 2\pi \delta(\omega) \quad . \quad . \quad . \quad . \quad . \quad (7b)$$

S. O. Rice<sup>(11)</sup> has also observed an equivalent result in the mathematical analysis of random noise.

Proceeding by the Fourier method,

$$\begin{aligned} i(t) &\equiv A(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{Y(\omega)}{j\omega} e^{j\omega t} d\omega + \frac{1}{2} \int_{-\infty}^{\infty} Y(\omega) \cdot \delta(\omega) \cdot e^{j\omega t} d\omega; \\ i. e. \quad A(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{Y(\omega)}{j\omega} e^{j\omega t} d\omega + \frac{1}{2} Y(0). \quad . \quad . \quad . \quad . \quad (8) \end{aligned}$$

A more elegant form can then be obtained by differentiating both sides with respect to  $t$  (assuming the validity of the step)

$$A'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) \cdot e^{j\omega t} d\omega, \quad . \quad . \quad . \quad . \quad . \quad (9)$$

which provides the desired relation. It follows, in addition, however, by Fourier's integral theorem that

$$\begin{aligned} Y(\omega) &= \int_{-\infty}^{\infty} A'(t) \cdot \epsilon^{-j\omega t} dt, \\ &= \int_0^{\infty} A'(t) \cdot \epsilon^{-j\omega t} dt. \quad \dots \dots \dots (10) \end{aligned}$$

[since  $A(t)=0$  ;  $t<0$ ].

This yields immediately the relation :

$$Y(0)=A(\infty). \quad \dots \dots \dots (10 a)$$

It should be noted here that A. W. Gent<sup>(12)</sup> derived a limited form of this latter result working *ab initio* by the Duhamel method in connection with pulse-distortion problems. Equations (9) and (10) then display readily the attractive reciprocity between  $Y(\omega)$  and  $A'(t)$ , which is fundamentally the basis of the pulse-stimulus method where the stimulus is a very short rectangular pulse. Alternatively, from (8), using the Fourier reciprocity,

$$Y(\omega)=j\omega \int_{-\infty}^{\infty} \left( A(t) - \frac{Y(0)}{2} \right) \cdot \epsilon^{-j\omega t} dt. \quad \dots \dots \dots (11)$$

Or, using (10 a) and remembering  $A(t)=0$  ;  $t<0$  :—

$$Y(\omega)=j\omega \int_0^{\infty} A(t)\epsilon^{-j\omega t} dt - \pi \cdot \delta(\omega) \cdot A(\infty). \quad \dots \dots \dots (11 a)$$

Thus from (10) or (10 a) one could determine  $Y(\omega)$  from a knowledge either of  $A(t)$  or  $A'(t)$ . In principle a "Fourier Transformer," such as that devised by Born, Fürth and Pringle<sup>(13), (14)</sup> might be used for the purpose.

The principle expressed in equations (9) and (10) has a perhaps more direct value in connections such as those suggested by Gent<sup>(12)</sup>, namely, the postulation of a suitable expression for the distortion tolerable in the response to a rectangular- (or under certain circumstances pulse-) stimulus — $A(t)$ —and the consequent derivation of an expression for the requisite network complex admittance ; this may then be approximated to by physical elements with a knowledge of the characteristics of various circuits.

Thus, for example, if a response composed of two parabolic arcs (see fig. 3) be tolerable, then we can derive the requisite  $Y(\omega)$  as follows :

$$\begin{aligned} A(t) &= 0 \quad (t < 0) \\ &= 2t^2 \quad (0 < t < \tfrac{1}{2}) \\ &= 1 - 2(1-t)^2 \quad (\tfrac{1}{2} < t < 1) \\ &= 1 \quad (t > 1). \\ \therefore A'(t) &= 0 \quad (t < 0 ; t > 1) \\ &= 4t \quad (0 < t < \tfrac{1}{2}) \\ &= 4(1-t) \quad (\tfrac{1}{2} < t < 1). \end{aligned}$$

$$\begin{aligned} \therefore Y(\omega) &= \int_0^{\frac{1}{2}} 4t e^{-j\omega t} dt + \int_{\frac{1}{2}}^1 4(1-t) \cdot e^{-j\omega t} dt \\ &= \frac{\sin^2 \frac{\omega}{4}}{\left(\frac{\omega}{4}\right)^2} \cdot e^{-2j\frac{\omega}{4}}, \text{ which is sketched in fig. 4.} \end{aligned}$$

Fig. 3.

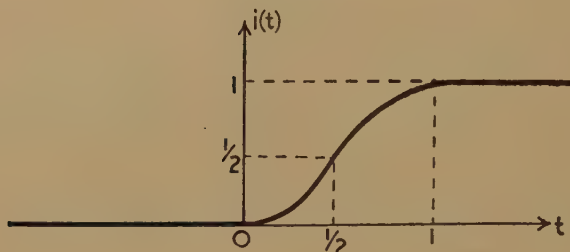
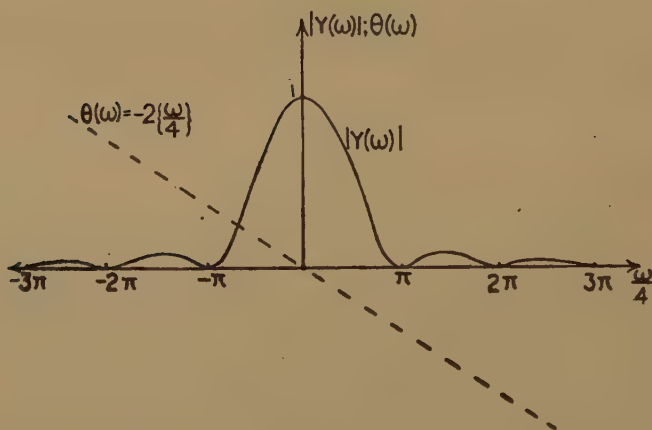


Fig. 4.



#### 4. "Pulse Response."

The specification of a network by its pulse response—*i. e.* ideally an infinitely short duration pulse of finite energy

$$\left( \frac{\partial}{\partial t} H(t) \equiv \delta(t) \right)$$

has been reviewed in a detailed manner in a recent paper by E. C. Cherry<sup>(15)</sup> and earlier writers have also considered the problem—in particular, the synthesizing of the response of a network by the method of "paired echoes."<sup>(16), (17)</sup> The theory is fundamentally closely related to that of the indicial admittance by the derivative relationship mentioned above, and



from many points of view undoubtedly affords considerable advantages. Alternative forms of pulse response might be considered, as, for example, has been done by Roberts and Simmonds <sup>(1)</sup>.

### 5. "Exponential Admittance (and Impedance)."

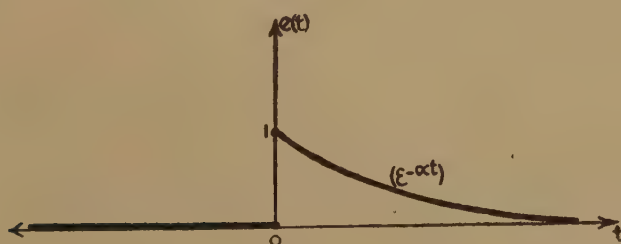
In view of the fact that, in principle, ideally perfect exponential stimuli of the form  $\epsilon^{-\alpha t} \cdot H(t)$  and  $(1 - \epsilon^{-\alpha t}) \cdot H(t)$  can be obtained easily in electrical circuits from familiar relaxation or "time-base" arrangements, it seemed of interest to the writer to consider whether the response to an applied signal of this form might prove a feasible specification. If, for example, a network is composed of a highly inductive element, then its indicial response will increase with time,

$$A(t) = \frac{1}{Lp} \cdot H(t) = \frac{t}{L} \cdot H(t),$$

ideally, and it will therefore require a "very long" duration rectangular-wave input before the network arrives at an approximately steady condition. Concomitantly it may prove difficult to design a generator to provide a reasonably "flat-topped" rectangular stimulus of the required duration. As another consideration it will be realized that physical networks composed of fundamental elements inevitably involving dissipation will display inherent decay properties, generally characterized by "time-constants." The application of a stimulus itself involving a time-constant of decay might therefore prove of value in the analysis of such networks analogously to the application of a signal characterized by a certain frequency to a network containing reactive elements and therefore displaying resonant properties.

Let us then consider a stimulus (see fig. 5) :

Fig. 5.



$$e(t) = 0 (t < 0) \\ = \epsilon^{-\alpha t} (t > 0),$$

$$\text{then } E(\omega) = \int_{-\infty}^{\infty} e(t) \epsilon^{-j\omega t} dt, \text{ or } E(p) = \int_0^{\infty} e(t) \epsilon^{-pt} dt \\ = \frac{1}{\alpha + j\omega} \quad \quad \quad = \frac{1}{\alpha + p}.$$

Therefore, the response or "exponential admittance" is given by

$$i(t) \equiv \mathbf{E}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{Y(\omega)}{\alpha + j\omega} e^{j\omega t} d\omega. \quad (12)$$

(which may be compared with (8)).

Thus by the integral theorem

$$\frac{Y(\omega)}{\alpha + j\omega} = \int_{-\infty}^{\infty} \mathbf{E}(t) \cdot e^{-j\omega t} dt,$$

or

$$Y(\omega) = (\alpha + j\omega) \int_0^{\infty} \mathbf{E}(t) \cdot e^{-j\omega t} dt \quad (13)$$

(since  $\epsilon(t) = 0$ ;  $t < 0$ ).

Alternatively, from (12),

$$\mathbf{E}'(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{Y(\omega) \cdot j\omega}{\alpha + j\omega} \cdot e^{j\omega t} d\omega.$$

$$\therefore Y(\omega) = \frac{\alpha + j\omega}{j\omega} \int_0^{\infty} \mathbf{E}'(t) \cdot e^{-j\omega t} dt \quad (14)$$

$$= \int_0^{\infty} \{\alpha \mathbf{E}(t) + \mathbf{E}'(t)\} e^{-j\omega t} dt. \quad (15)$$

This latter result indicates, then, how the complex admittance may be synthesized from the exponential admittance as from the indicial admittance earlier and also by direct comparison with (10) the relationship between the indicial admittance and the exponential admittance follows, namely :

$$\underline{A'(t) = \alpha \mathbf{E}(t) + \mathbf{E}'(t) = (\alpha + D)\mathbf{E}(t).} \quad (16)$$

Finally, we can therefore show by substitution in the Duhamel integral, that the response to *any* stimulus in terms of the exponential admittance is

$$i(t) = \int_0^t \{(\alpha + D) \cdot \mathbf{E}(t - \lambda)\} e(\lambda) \cdot d\lambda. \quad (17)$$

The three fundamental elements R, L, and C have thus admittances :

	$\mathbf{E}(t)$	$\mathbf{E}'(t)$	$A(t)$	$A'(t)$
R	$\frac{1}{R} e^{-\alpha t}$	$-\frac{\alpha}{R} e^{-\alpha t}$	$\frac{1}{R}$	0
L	$\frac{1}{\alpha L} (1 - e^{-\alpha t})$	$\frac{1}{L} e^{-\alpha t}$	$\frac{t}{L}$	$\frac{1}{L}$
C	$-\alpha C e^{-\alpha t}$	$\alpha^2 C e^{-\alpha t}$	0	0

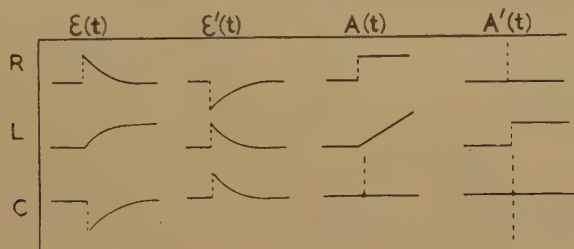
(It should, of course, be noted that in the case of C we exclude the point  $t=0$ —otherwise a Dirac impulse function must be included.)

These admittances are sketched in fig. 6 and it appears that the forms of the exponential admittance are at least as recognizable for the fundamental elements as those of the indicial admittance. Examining this table, it is evident that a combination of  $R$  and  $L$  in parallel, such that  $\frac{R}{L} = \alpha$  provides as exponential admittance simply  $\frac{1}{R}$ . Thus the "analogue" of a pure resistance in indicial admittance is a parallel combination of  $R$  and  $L$  in exponential admittances to an index  $\alpha = \frac{R}{L}$ .

The vital feature then of an "exponential signal generator" would be the parameter  $\alpha$ , which is the direct analogue of angular frequency  $\omega$ , in a (frequency) signal generator. It is suggested that this parameter  $\alpha$  would be calibrated directly in terms of the decay constant of the output.

From an exponential signal generator it should be relatively straightforward to provide four outputs in all—two complementary wave forms for an impressed voltage stimulus, and two for an impressed current

Fig. 6.



stimulus: namely,  $e^{-\alpha t}$  and  $1 - e^{-\alpha t}$ . The admittance current responses to these voltage stimuli let us call  $E(t)$  and  $E_o(t)$ , and the impedance (voltage) responses to the current stimuli  $\zeta(t)$  and  $\zeta_o(t)$ . It is seen readily that

$$E_o(t) = \alpha \int_0^t E(t) dt; \quad \zeta_o(t) = \alpha \int_0^t (\zeta) dt.$$

A table can then be set up for the four "fundamental" elements as in fig. 7. Thus each "element" has a corresponding immediately recognizable exponential admittance or impedance (those underlined) when  $\alpha$  is correctly adjusted and behaves than as a pure, idealized resistance under indicial conditions; the other responses are, however, not without interest; for example, the second element has  $\zeta_o(t)$  which behaves as the admittance of a pure inductive element under indicial conditions, and other responses correspond to those of a "critically damped" tuned circuit under indicial conditions. It is also of interest to examine the response of the "elements" to applied exponential signals such that

$$\alpha \neq \frac{R}{L} \quad \text{or} \quad \frac{1}{CR}.$$

Considering the first "element" as typical, it may be seen that it naturally belongs to the "admittance" class. (No  $e^{-\alpha t}$  factor appears in

either  $\epsilon(t)$  or  $\epsilon_e(t)$ , and similarly with the other elements.) If now we apply exponential stimuli to index  $\alpha$ , where, say,  $\frac{R}{L} = \beta$ , we find

$$\epsilon(t) = \frac{\beta}{\alpha} \left\{ \frac{1}{R} \left( 1 + \frac{\alpha - \beta}{\beta} e^{-\alpha t} \right) \right\} \dots \dots \dots (18)$$

and

$$\epsilon_e(t) = \frac{\beta}{\alpha} \left\{ \frac{1}{R} \left( \alpha t + \frac{\alpha - \beta}{\beta} [1 - e^{-\alpha t}] \right) \right\}, \dots \dots \dots (19)$$

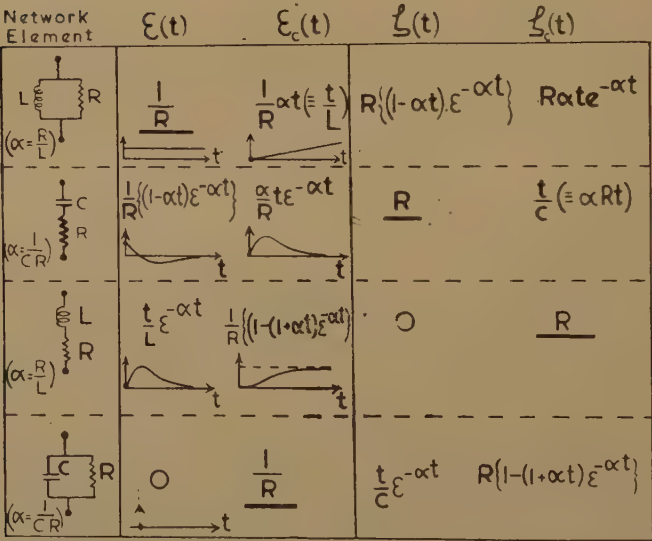
while, for example, if we consider the last network, we find

$$\epsilon(t) = \frac{\alpha}{\beta} \left\{ \frac{1}{R} \left[ 0 + \frac{\beta - \alpha}{\alpha} e^{-\alpha t} \right] \right\} \dots \dots \dots (20)$$

And similarly,

$$\epsilon_e(t) = \frac{\alpha}{\beta} \left\{ \frac{1}{R} \left( 1 + \frac{\beta - \alpha}{\alpha} (1 - e^{-\alpha t}) \right) \right\} \dots \dots \dots (21)$$

Fig. 7.



Similar equations naturally hold for  $\zeta(t)$  and  $\zeta_e(t)$  for the remaining two "elements." These solutions then indicate that for the basic elements we may regard the behaviour of their appropriate admittances (or impedances) to any index  $\alpha$ , as a "steady-state" solution (the first term) corresponding to their behaviour when an exponential signal of the characteristic index is applied plus a "transient" term due to the signal index which vanishes if the characteristic index is employed. Such behaviour may be seen to have a ready analogue for circuits in which reactive properties are of primary consideration under the stimulus of a sinusoidal signal generator.



### 6. Present Conclusion.

Although the suggested theory of application has been investigated in only a very preliminary survey, the writer feels that an exponential signal of variable index might form a useful aid in network analysis and that an exponential "signal generator" might well form a fundamental instrument analogous to a "frequency" signal generator. Two general aims remain : to investigate the theory further with reference to general networks composed of (inevitably) "lossy" elements and to carry out practical work to extend and verify the findings.

### References

- (1) F. F. Roberts and J. C. Simmonds, *Phil. Mag.* (Dec. 1943, July 1944).
- (2) Bloch, A., *Jour. Inst. Elec. Eng.* xci. (pt. 3), p. 31 (1944).
- (3) McLachlan, N. W., 'Complex Variable and Operational Calculus.' (Cambridge University Press, 1939.)
- (4) Gardner & Barnes, 'Transients in Linear Systems.' (Wiley, New York, 1942.)
- (5) Starr, A. T., 'Electric Circuits and Wave Filters' (particularly cap. XII. for Fourier Theory). (Pitman, 1938, ed.2.)
- (6) Shea, T. E., 'Transmission Networks and Wave Filters' (particularly cap. XI.) (Van Nostrand, 1929),
- (7) Bush, V., 'Operational Circuit Analysis.' (Wiley, New York, 1929.)
- (8) MacDonald, D. K. C., "Network Analysis Involving Realizable Filter Functions," *Phil. Mag.* (in press).
- (9) Bush, V., *loc. cit.* cap. IV.
- (10) Bush, V., *loc. cit.* cap. V. (The theorem is alternatively known as the "superposition theorem.")
- (11) Rice, S. O., *Bell. Syst. Tech. Jour.* xxiii. p. 282 (1944) ; xxiv. p. 46 (1945).
- (12) Gent, A. W. Unpublished memorandum.
- (13) Fürth, R., & Pringle, R. W., *Phil. Mag.* xxxv. p. 643 (1944).
- (14) Born, M., Fürth, R., & Pringle, R. W., 'Nature,' clvi. p. 756 (1945).
- (15) Cherry, E. C., *Jour. Inst. Elect. Eng.* xcii. (pt. 3), p. 183 (1944).
- (16) Wheeler, H. A., *Proc. Inst. Rad. Eng.* xxvii. p. 359 (1939).
- (17) Zworykin, V. K., & Morton, G. A., 'Television,' p. 420 *et seq.* (Short, but useful, note illustrating the application of the method.) Wiley, New York, 1940.)

---

### LXXXI. *The Luminous Efficiency of the High Pressure Mercury Vapour Discharge* \*.

By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., G. G. ISAACS, B.Sc.,  
A.Inst.P., A.M.I.E.E., and E. H. NELSON.

[Received August 23, 1946.]

### SUMMARY.

Experimental evidence is produced to show it is likely that the maximum luminous efficiency obtainable with the high pressure mercury

---

\* Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England.

vapour discharge is in the neighbourhood of 85 lumens per watt. This is consistent with what would be expected from fundamental considerations.

### *Introduction.*

THE high pressure mercury vapour lamp<sup>(1)</sup> was one of the first lamps in which the high luminous efficiencies obtainable by means of electric discharges through gases was made use of for general lighting purposes. The high pressure mercury vapour lamps, together with sodium lamps, began to come into use for street lighting about 1932. The potentialities of the sodium vapour discharge are fairly well understood<sup>(2)</sup>, but so far there has been no authoritative estimate of what might be obtainable from the mercury vapour discharge. For example, the work of Marden, Meister and Beese<sup>(3)</sup> indicates that efficiencies greater than 60 lumens per watt may be obtainable, but they give no clear indication of where and in what circumstances an upper limit may be expected.

The mercury lamps at present on the market have luminous efficiencies of 40–45 lumens per watt, and it is generally understood that somewhat higher efficiencies are obtainable in units of larger wattage rating. For the water-cooled lamp, which makes use of a discharge through a capillary with a mercury vapour pressure of the order of 100 atmospheres, a figure of 60 lumens per watt has been quoted<sup>(4)</sup>. During the course of the authors' work experiments have shown that the luminous efficiencies obtainable are in fact higher than have so far been suggested; and, as the results have implications of definite theoretical interest, the present note has been thought worth while.

There are two efficiency figures which have to be considered: the intrinsic efficiency of the arc, which is of the greater interest from the academic point of view, and the overall efficiency of the discharge tube, which is lower than the intrinsic efficiency as a result of electrode and other losses. So far as the luminous efficiency of the arc itself is concerned there are two main factors which are known to be of major importance: the watts per centimetre of the discharge column and the vapour pressure or, what is equivalent, the voltage drop per centimetre of the column. A less important factor is the geometry of the envelope; this influences the luminous efficiency by the way in which it may contribute to the cooling of the arc, particularly in respect to the manner in which convection losses are involved. It is not likely, however, that, within limits of practicability, the geometry of the envelope is of major importance.

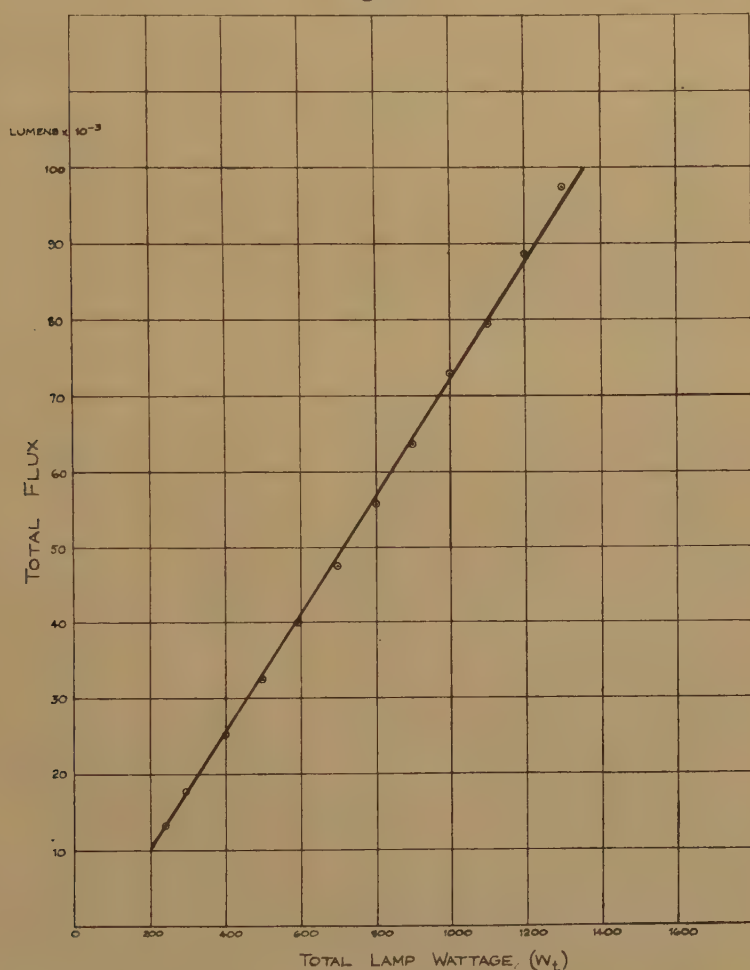
The present commercial lamps used for general lighting purposes operate with a loading of some 20–50 watts per centimetre and a voltage drop of 10–40 volts per centimetre of arc. In this region it is known that the luminous efficiency increases with increasing loading, and it is also known that the efficiency falls off if the voltage drop is decreased

very much. The present experiments were therefore confined to the region of operation in which high luminous efficiencies might be expected, namely high wattage loading and relatively high pressures.

*Experimental.*

The discharge tubes used all had quartz envelopes of cylindrical form, operated with the arc vertically along the axis. The internal diameters

Fig. 1 a.

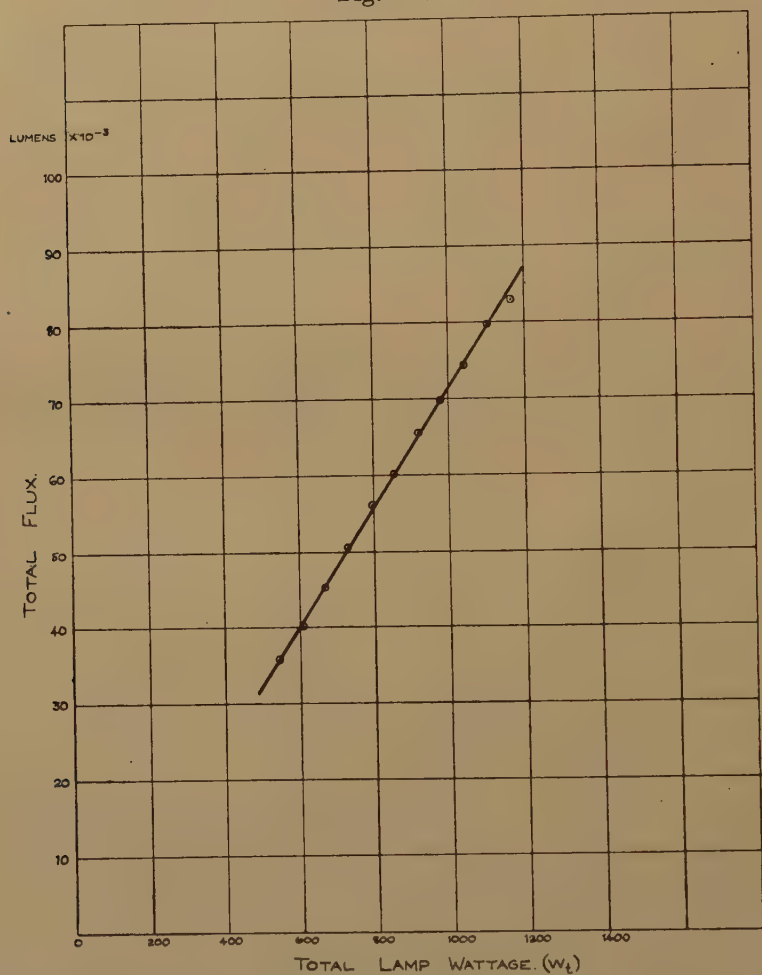


Arc length 15 mm., internal diameter 30 mm., tube voltage 130 volts.

ranged between 25 and 30 mm. (with one exception—20 mm.). Electrode constructions did not vary appreciably and consisted essentially of tungsten rods about 4 mm. diameter tapered at the tips and carrying activating material for starting and initial running conditions. The quartz-molybdenum seals were of comparatively small diameter (approximately 10 mm.) so that light obscuration was minimised.

Measurements were made on both A.C., with ordinary inductor stabilizing, and D.C. There was no significant difference of lumen output under the conditions in which the experiments were carried out. It is, of course, known that this is not universally true, but it is to be expected in the case of the high pressure mercury discharge at high loadings.

Fig. 1b.



Arc length 50 mm., internal diameter 20 mm., tube voltage 400 v.

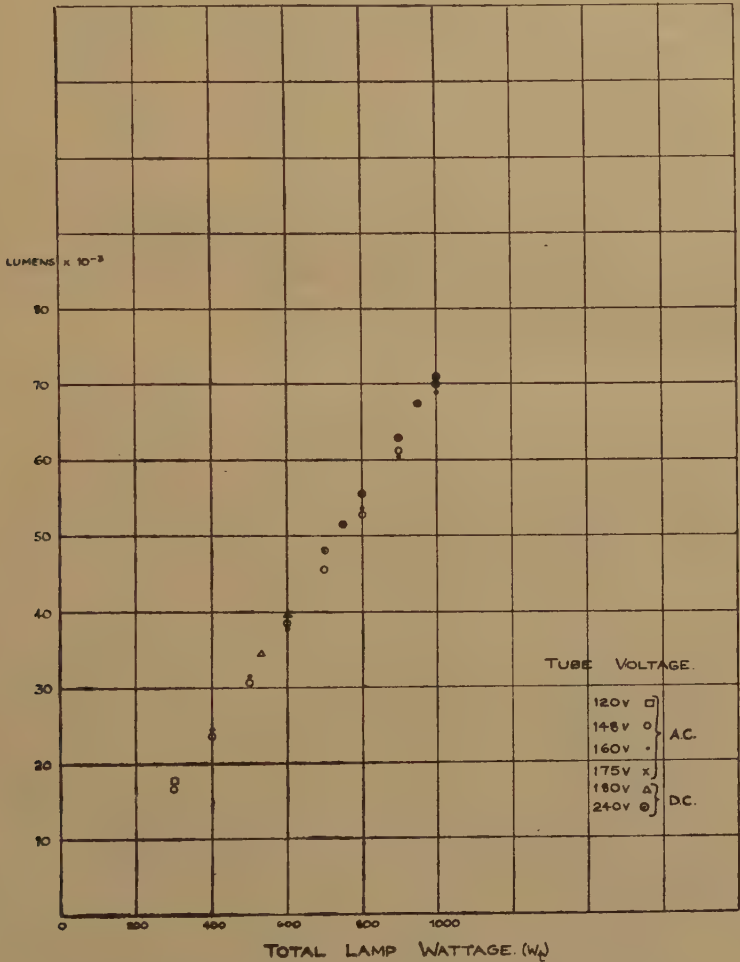
Measurements of luminous flux (L) were made on some 10 tubes, over a range of input wattages for which the voltage across the tube did not change appreciably. A limit to the range was given when the input was insufficient to maintain the mercury vapour pressure; but a wide enough range could be covered without taking measurements at very low wattage inputs where it would be necessary to apply external heating in order to prevent a drop of arc voltage. In order to facilitate rapid



measurement, the luminous flux readings were first recorded simply as photo-cell currents which were afterwards calibrated by careful photometry at one value of wattage ( $W_t$ ) and voltage ( $V_t$ ) in the set of readings for each experimental tube.

Examples of the lumen-wattage curves obtained are shown in figs. 1 a, 1 b. It will be noted that, over the range considered, they

Fig. 1 c.



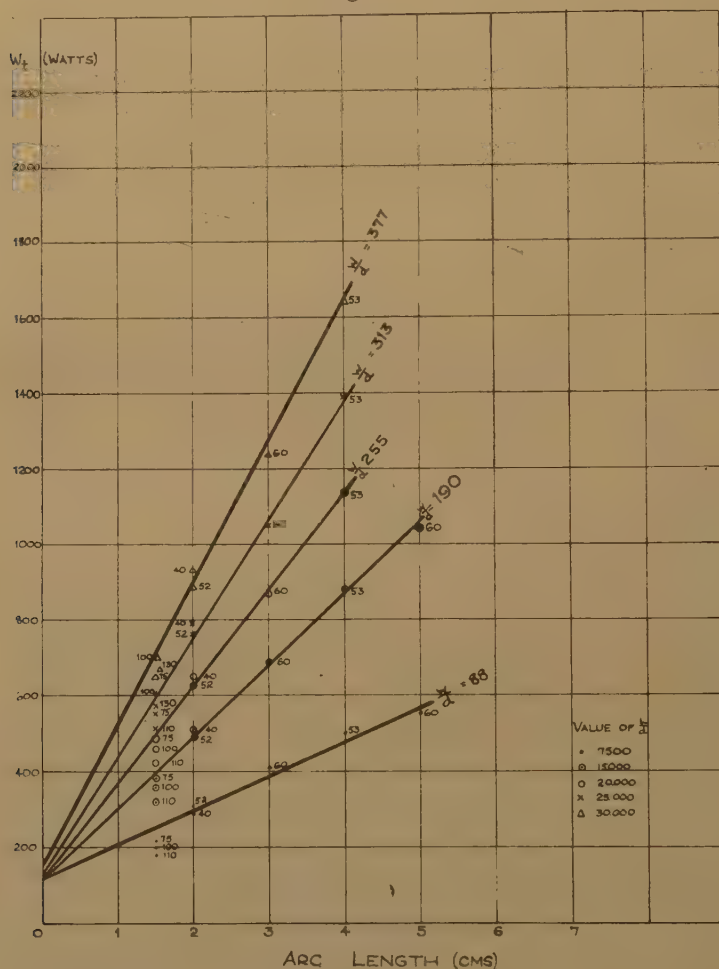
Arc lengths all 15 mm., internal diameters all 25 mm.

approximate closely to straight lines. It became evident as measurements accumulated that the efficiency was not very dependent on the arc voltage. Fig. 1 c, for example, shows the measurements taken on five tubes over a fairly wide range of voltage drop.

These results and others of the same type indicated that over the watts-per-cm. range at which the experiments were made a 1 per cent. change of lumen output resulted from 25 per cent. change of volts. This

is only a rough estimate because the data were insufficient to provide a more accurate estimate of a small dependence such as this; and moreover a more accurate estimate was not necessary for our purpose. The arc voltage drop could not easily be increased above the maximum values given here, as the strength even of quartz envelopes was inadequate to withstand the high internal pressures developed; the consequent limitation in the results is discussed later.

Fig. 2.



The results shown in fig. 1 and other curves of a similar type were used to determine the lumen-watt characteristics of the arc itself (lumens per watt against wattage per centimetre of arc length) by separating the total input wattage ( $W_t$ ) into two parts, one of which ( $W$ ) is dissipated in the arc itself and contributes directly to light production while the other ( $w$ ) is dissipated at the electrodes. The method adopted was as follows:—

Using all the available readings, but without extrapolation, the sets

of points in fig. 2 (and other similar sets) were plotted. Each set is for constant lumens per cm. ( $L/d$ ), and shows the relation between arc length ( $d$ ) and total wattage  $W_t$ . The values of  $L/d$  range between 5,000 and 35,000 lumens per cm. It should be noted that, from the numbers marked against the points (which are the approximate values of the voltage drop per centimetre of arc), the indication is again that this is not a major factor. If the relations between  $W_t$  and  $d$  were, in fact, found to be accurately linear, the intercept of the straight lines on the  $W_t$  axis would give the electrode loss. Deviations from the linear relation are due to several factors; the most important probably are: firstly that the electrode loss depends to a greater extent on the arc voltage drop than does the luminous flux, and secondly that variations in these comparatively small losses must occur even in tubes of the same apparent design. Since these losses are relatively small these differences will not seriously affect the curves shown in fig. 1, but they are quite sufficient to prevent the ideal straight line being obtained in fig. 2. A further consideration is that whereas any error in calibrating the photocell readings will not appear at all in fig. 1, such errors will make themselves felt as soon as results from more than one set of photocell readings are used together.

Nevertheless, the relation between  $W_t$  and  $d$  is sufficiently nearly linear to make it justifiable to obtain a first approximation to the electrode losses. If, as appears to be the case, these losses are relatively small, a first approximation is all that is necessary for our purpose. The points corresponding to  $d=1.5$  are of rather higher arc voltage gradient than the others. This fact was taken into account, and an allowance made for it. With this adjustment for arc voltage the estimated best straight lines through each set of points for constant  $L/d$  were drawn.

Accepting the intercept on the  $W_t$  axis of each of these lines as representing the end losses for that particular value of  $L/d$  for which the line is drawn, it follows that the lumen output values for any particular value of  $W/d$  can be immediately obtained. In particular, the slope of these lines allows the required value of  $L/W$  for each value of  $L/d$  to be found and therefore for each value of  $W/d$ .

This relation between  $L/W$  and  $W/d$  is shown in fig. 3 as the full line through the points obtained in the manner indicated. It will be noted that a point at  $W/d=88$  has been neglected, due to the uncertainty of the slope of the corresponding line in fig. 2.

The purpose which prompted this work did not require the extension of the curves to lower values. It is, however, fairly clear from known data on lamps of the same type as the present commercial lamps that at about 25 watts per cm. the corresponding efficiency figure is something above 50  $L/W$ ; the curve, therefore, will probably tail-off approximately as shown.

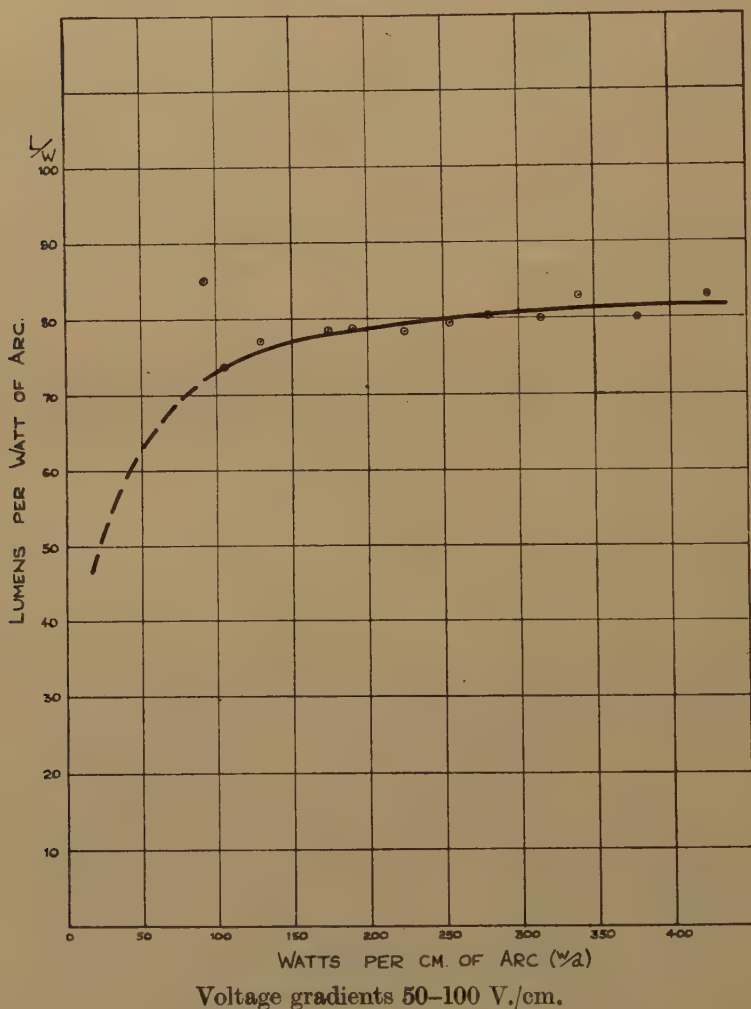
#### *Discussion of Results.*

Fig. 3 gives a clear indication that the efficiency curve is increasing very slowly, if at all, with watts per cm. at the higher values. The

indication also is, from the data shown in fig. 1 c, that if this curve could be obtained for much higher pressures, it would still not be appreciably higher. There does therefore seem to be a clear suggestion that the maximum efficiency is likely to be near 85 L/W.

There is an interesting point of theoretical significance involved here. If the mercury vapour pressure were increased, conditions would be

Fig. 3.



approached wherein the radiation from the discharge would approximate to that of a black body. The amount of continuous radiation present in the tubes used in these experiments is such as to suggest that the volts per cm. would not need to be increased by a very large factor before an approximation to this state of affairs is reached. There is also additional evidence from the water-cooled capillary lamp that at gradients of the



order of 1,000 V./cm. the radiation approaches black body condition <sup>(5)</sup>. Now the black body at the optimum temperature of about 6,500° K. gives a luminous efficiency of about 84 L/W. There is theoretical evidence <sup>(6)</sup> which shows that the discharge temperature will be in the neighbourhood of 6,000° K. in the region between that for which results are given in this paper and the region where substantially continuous radiation might be expected. If the discharge temperature in the latter condition is not the optimum, the efficiency will be below 84 L/W, and therefore for any particular  $W/d$  we might expect the efficiency to decrease with increasing  $V/d$  for some large value of  $V/d$ . There will, however, be some range of values of  $W/d$  for which substantially the optimum temperature is obtained and therefore the optimum efficiency of 84 L/W. The question therefore arises whether, for any value of  $W/d$  on the curve of fig. 3 if the pressure and  $V/d$  are steadily increased, the efficiency approaches asymptotically the value of 84 L/W. It is, of course, possible that due to the line radiation the efficiency may go higher than 84 L/W before the black body radiation limitation takes control, but this does not appear very likely on general grounds. The most likely interpretation of these results, therefore, appears to be that the whole of the curve of fig. 3 lies below 84 L/W and that the part of it for which the temperature conditions happen to be right will approach the 84 L/W asymptotically as the  $V/d$  increase. For other parts of the curve (or its extension to higher values of  $W/d$ ) the optimum efficiency as a function of  $V/d$  may be less than 84 L/W.

#### References.

- (1) Ryde, G. E. C. *Journal*, iv. p. 199 (Nov. 1933).
- (2) Pirani, *Z. tech. Phys.* xi. p. 483 (1930); Krefft, Reger and Rompe, *Z. tech. Phys.* xiv. p. 242 (1933); Karfeld and Taraskov, *Tech. Phys. U.S.S.R.* iii. p. 881 (1936).
- (3) Marden, Meister and Beese, *Transactions of the Electrochemical Society*, lxi. p. 389 (1936).
- (4) Dorgelo, *Philips Tech. Rev.* p. 165 (1937).
- (5) Elenbaas, *Physica*, iii. p. 859 (1936).
- (6) Francis, *Phil. Mag.* (7) xxxvii. p. 433 (1946); *ibid.* p. 653 (1946).

#### LXXXII. On the Outer Unipolar Rotation and the Outer Unipolar Induction.

By MARIN KATALINIĆ\*.

[Received May 9, 1946.]

1. FARADAY'S well-known experiment of rotating a wire carrying current around the end of a stationary magnet bar is often considered as an inverse of the unipolar rotation (Ampère's experiment). An application of the action and reaction principle to such a relation between

\* Communicated by the Author.

both phenomena can lead to incorrect rules regarding the sense of rotation\*. As a matter of fact, Faraday's experiment is identical with the unipolar rotation experiment; respectively it is a special case of the outer unipolar rotation, where a conducting wire, which may or may not be connected with the magnet bar, rotates under the influence of the outer magnetic flux of the same magnet. The efficacious torque is therefore in its sense and in its amount equal to the torque given by the formula of Diesselhorst† or of Zeleny and Page‡ for the inner unipolar rotation, if for  $\Phi$  is taken the outer magnetic flux between the upper end of the rotating wire and the sliding contact, cut by the wire in a revolution. On the other side the equivalent torque of the inner unipolar rotation results, if for  $\Phi$  one takes the inner magnetic flux of the magnet, cut in its revolution, limited between its axis and the flux line passing through the same sliding contact. It is useful to imagine the elements of the movable wire carrying the current projected along the corresponding flux lines of the magnet into its outer equatorial plane. As it is the property of the equator of a magnet to be for any path of integration the like pole of the electromotive forces induced by rotating an outer wire or the magnet respectively in the same sense, so also here, on account of the energy law, or with regard to the signs of the Lorentzian forces, the sign of the electrical current producing a unipolar rotation of the magnet or of an outer wire in the same sense has to be defined with regard to the equator of the magnet. Therefore, the same sign belongs to two currents flowing in an outer wire or in the magnet towards its equator\*, whereas a current flowing in a side arm to the equator has to change its sign entering the magnet. In this manner, for both phenomena, the unipolar induction and the unipolar rotation, the opposite sense of the magnetic flux outside of the magnet is in an equal way taken into account.

A tripartite, symmetrically magnetized, cylindrical magnet of cobalt steel ( $l=22.2$  cm.,  $2r=4.0$  cm.,  $\Phi=2,160 \cdot 10^4$  maxwells) consisting of three axially hollow magnets, which were insulated from each other†, (fig. 1), was suspended by means of a steel wire ( $l=41.5$  cm.,  $2\rho=0.11$  mm.). The deflections of the magnet were measured by means of a mirror  $M$ , attached to the suspension, and a scale. To assure definite deflections, the upper and the lower end of the wire were jammed in special brass clamps ( $b$ ) of the suspension device. To the lower end of the magnet was screwed an ebonite plate, which carried the support of an

\* See, for example, O. D. Chwolson, *Lehrb. d. Phys.* iv./1, Braunschweig, p. 772 (1908).

† H. Diesselhorst, *Elektrodynamik* (Graetz, *Handb. d. Elektr. u. d. Magnet.* iv. Leipzig, p. 1121–1326, 1920), p. 1323–1326.

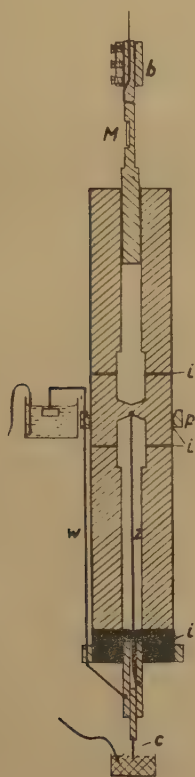
‡ J. Zeleny and L. Page, *Phys. Rev.* (2) xxiv. p. 544 (1924); xxvii. p. 470 (1926); xl. p. 299 (1932).

\* Both currents have in the suspension wire the same sense.

† This magnet served previously to other investigations. In the schematic fig. 1 the details of its construction are omitted;  $i$  denote the insulations. Needless to say, that in all experiments here described all iron was removed from the neighbourhood of the magnet.

axial platinum contact (*c*) dipping in mercury. To the middle part of the compound magnet was fixed a brass ring (*p*), insulated from the magnet body by celluloid. The upper face of this ring was lying in the equatorial plane; in the inner side of this face a shallow groove was engraved in order to make contact with the magnet body by means of a mercury drop. The ring carried a short conducting side arm terminated by a platinum electrode dipping in an electrolyte solution. Moreover, the ring had 3 mm. off the magnet surface an axial hole, where a lateral wire (*w*) could be introduced. This wire went downwards along the magnet mantle, its lower end being fixed to the support of the axial platinum

Fig. 1.



contact. If this lateral wire was not put in, and the groove on the upper face of the ring (*p*) contained mercury, normal inner unipolar rotation was realized by the current flowing from the electrolyte solution through the side arm to the middle part of the magnet, then coming out through the axial wire (*z*) and the axial contact (*c*). By introducing the lateral wire (*w*) in the hole and removing the mercury from the groove, one had an immediate transition to the outer unipolar rotation due to the current passing through the lateral wire. Obviously the latter arrangement coincides fully with Faraday's experiment, the circumstance, that the lateral wire was here fixed to the support of the axial platinum

The oscillations of the deflected magnet were almost undamped. A complete quieting of the magnet in this and in the following experiment was easily obtained by a fine counter-regulating of the current; *i. e.*, the current intensity was made slightly increasing, when the elongations decreased, and contrariwise, in the rhythm of the oscillations. The north end of the magnet was uppermost. The small torque arising from the earth's magnetic field, which acts equally in both cases, was neglected. Eventual constructive asymmetries were neutralized by reversing the sense of the current.

By a current intensity  $i=100.0$  mA. the following deflections of the magnet were obtained:—

For the inner unipolar rotation:  $+5.82^\circ \pm 0.005^\circ$ .

For the outer unipolar rotation:  $+5.85^\circ \pm 0.008^\circ$ .

The slight difference of  $0.03^\circ$  is prevailingly real. It corresponds to the negative torque generated in the first case by the current flowing in the innermost 3 mm. of the brass ring close to the magnet (between the magnet surface and the axial hole for the lateral wire), the negative torque in the side arm and in the remaining part of the ring being in both cases equal.

2. Since the fundamental paper of Lenz\* the unipolar induction and the unipolar rotation have been considered as inverse phenomena. Hagenbach†, Diesselhorst‡, Zeleny and Page§ derived this reciprocity from the energy law. The ultimate cause of this reciprocity lies in the fact that in both phenomena the Lorentzian forces acting upon the free electrons are effective. A quantitative experimental proof of this reciprocity in the case of the inner unipolar rotation and of the inner unipolar induction can be obtained only by absolute measurements of the torque  $L$  and of the integral induced electromotive force  $E$ , for the same value of the magnetic flux being

$$L = \frac{i}{\omega} \cdot E,$$

where  $i$  and  $\omega$  denote the current intensity in the unipolar rotation experiment and the angular velocity of the spinning magnet.

In an outer unipolar rotation experiment and in an outer unipolar induction experiment all points of the movable part of the circuit are accessible by suitable measuring devices. Thus the distribution of the total torques generated by a given current intensity over a movable wire carrying the current, respectively the distribution of the integrated electromotive forces induced by revolving the same wire, can be obtained by measurements. Therefore the outer unipolar rotation and the outer

\* E. Lenz, *Pogg. Ann.* xxxi. pp. 483, 493–494 (1834).

† E. Hagenbach, *Ann. d. Phys.* (4) iv. p. 233 (1901).

‡ Diesselhorst, *l. c.*, p. 1326.

§ Zeleny and Page, *l. c.*, xxiv. pp. 544, 550 (1924).



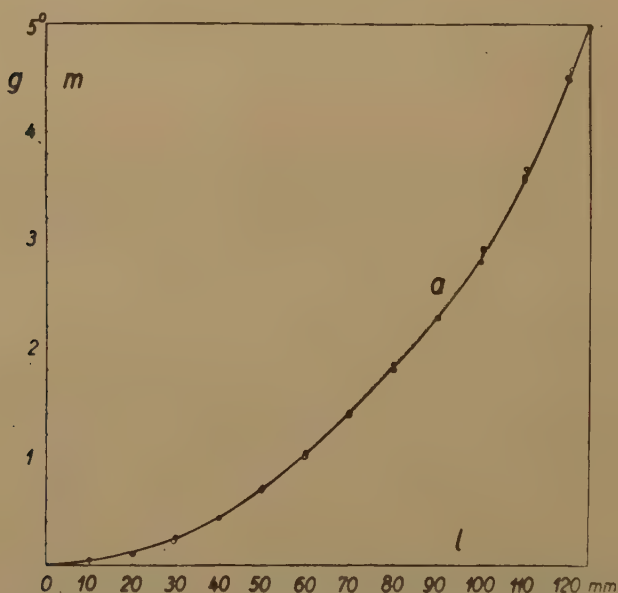
unipolar induction offer a possibility to obtain experimental evidence for the reciprocity between both phenomena by only comparative measurements.

In this regard, very simple is the case of the outer unipolar rotation produced by a current sent through a straight wire stretched along the mantle of a cylindrical magnet parallel to its axis, and of the outer unipolar induction produced by revolving the same wire around the axis of the magnet. In this case the generated torque is due to the Lorentzian forces acting only on the tips of arms of a constant length. Therefore, the distribution of torques over such a wire is reduced to the distribution of forces. The Lorentzian forces which generate the electromotive force induced by revolving the same wire, being to these forces proportional, it must be expected that in this case the curve representing the distribution of the total torques over such a wire should be similar to the curve representing the distribution of the integrated induced electromotive forces over the same wire; respectively, it should be possible to bring both curves to coincidence by multiplying the one with a suitable factor proportional to  $\frac{i}{\omega}$ . Of course, the same should be valid for these distributions over a meridian of the mantle of a cylindrical magnet.

The arrangements to test the principle of these conclusions were built up on a full cylindrical magnet of cobalt steel ( $l=25$  cm.,  $2r=4.0$  cm.,  $\Phi=2.749 \cdot 10^4$  maxwells). A brass contact ring, 5 mm. thick, was fixed to the equator of the magnet mounted in the rotatory apparatus, the axis of the magnet being perpendicular to the plane of earth's magnetic meridian. Another brass ring precisely as thick was movable along the magnet mantle: its distances from the magnet equator could be read on a scale drawn on the mantle. A narrow brass strip as a sliding contact slid on the equatorial ring exactly in the equator plane, while another narrow brass strip slid on the movable ring just at the desired distance ( $l$ ) from the equator. As the integrated induced electromotive force depends only upon the sites of the sliding contacts and is independent of the integrating path between them, the induced electromotive force was the same as it would be if an even wire parallel to the magnet axis rotated coaxial with it at a distance of 5 mm. from the magnet surface. By fixing the movable ring at various distances from the equator, the distribution of the induced integrated electromotive forces over such a revolving wire was obtained in this manner for a distance of 5 mm. from the magnet surface. For the purpose of the subsequent comparison with the deflections of the magnet, the deflections of the galvanometer were expressed in angular measure. The deflections coming from the parasitic thermo-electromotive forces were eliminated by reversing the sense of rotation. The fluctuations of the resistances in the sliding contacts were neutralized by a sufficient resistance (400 B.A.U.) in the galvanometer circuit. The distribution obtained is given in the curve *a* of fig. 2, the full circles representing the galvanometer deflections for a frequency of rotation of 1.0 turns/sec.

Then the magnet was suspended in the above-described manner. An insulated brass ring 5 mm. thick was fixed to it a little above its equator. To the one side of this ring was soldered a short thin platinum wire, whose lower end, lying exactly in the plane of the magnet equator 5 mm. from its surface, was slightly dipping in distilled mercury contained in a stationary wooden groove. To the opposite side of the brass ring was jammed a straight copper wire terminating at its lower end in a thin platinum wire, whose tip dipped slightly in distilled mercury, contained in another stationary wooden groove, at a distance of 5 mm. from the magnet surface. The copper wire and the groove could be lowered along the magnet mantle. The current was flowing from one groove with mercury through the brass ring and the copper wire to the mercury in the

Fig. 2.



other groove. In order to maintain the value of the damping factor nearly constant (1.04 to 1.05), the mercury in the grooves was renewed for any new measurement. The torques were measured by the magnet deflections expressed in angular measure. In this way the distribution of the total torques over this wire for the same distance from the magnet surface as above was obtained. The current intensity was 100.0 mA. The small torque due to the horizontal component of earth's magnetic field acting upon the copper wire was neglected.

As a basic value for the mutual comparison of the unipolar induction with the unipolar rotation, was chosen the average galvanometer deflection in the unipolar induction experiment for the distance  $l=12.5$  cm. of the movable sliding contact from the magnet equator. This value was  $g_0=4.980^\circ$ . The average deflection of the magnet ( $m$ ) by the unipolar

rotation for the same distance of the movable platinum contact from the magnet equator was  $m=5.724^\circ$ . This deflection was reduced to  $m_0=g_0=4.980^\circ$  and in the same ratio  $\frac{m_0}{m}$  were reduced the magnet deflections for other distances ( $l$ ) of the movable platinum contact from the equator. These reduced magnet deflections are shown by the empty circles on and around the path of the curve  $a$  in fig. 2. Thus, the curves representing the distributions of the integrated induced electromotive forces and of the generated torques coincide in this case with a great accuracy.

### *Conclusion.*

Faraday's experiment of revolving a wire carrying current around the end of a stationary magnet bar is a special case of the outer unipolar rotation and is quantitatively identical with the inner unipolar rotation (Ampère's experiment).

If the outer unipolar rotation is produced by a current flowing in a wire stretched parallel to the magnet axis, the torque is generated by Lorentzian forces having a constant arm. Therefore, the distribution of the total torques over such a wire represents the distribution of resultants of Lorentzian forces and coincides, by a suitable reduction, with the distribution of the integrated electromotive forces induced by revolving the same wire around the magnet. This coincidence constitutes a quantitative evidence of the reciprocity between the unipolar induction and the unipolar rotation.

The University,  
Physics Institute,  
Zagreb (Yugo-Slavia).  
September, 1942.

---

---

### LXXXIII. *The Thermal Ohm, Farad and Henry.*

By R. C. L. BOSWORTH, Ph.D., D.Sc., F.Inst.P.\*

[Received August 7, 1946.]

### *Introduction.*

THE analogy between heat flow and electric current has proved useful both as an educational aid and, in the hands of Paschkis (1942, 1946) and his co-workers, as a research tool for studying thermal transients by the method of electric models. The definition of the thermal potential difference or driving force as a temperature difference and the thermal current as the rate of heat flow have led naturally to the concept of thermal resistance. Harper (1928) has suggested the name "fourier" as

---

\* Communicated by the Author.

the metric unit of thermal resistance, while White (1938) has defined the thermal ohm as that unit of thermal resistance which allows 1 watt of heat energy to flow every second under a potential difference of  $1^\circ\text{C}$ .

Let  $\theta$  represent the temperature difference,  $q$  the emittance or units of heat (watts) crossing 1 sq. cm. in 1 second and  $a$  the cross-sectional area. Then, according to the definitions of White, we may write :

$$\text{Thermal p.d.} \quad V_q = \theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\text{Thermal current} \quad I_q = qa. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\text{Thermal resistance} \quad R_q = \frac{\theta}{qa}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Let  $V_e$ ,  $R_e$  and  $I_e$  represent the corresponding electrical quantities. With the definitions as given by equations (1), (2) and (3) the Fourier law for the conduction of heat becomes analogous to Ohm's law for the conduction of electric charge. However, when we attempt to extend this analogy to cover definitions of a thermal "farad" and a thermal "henry," difficulties are encountered. These difficulties are connected with the measure of the energy stored in the thermal field.

#### *The Energy Difficulty.*

If an electric charge  $Q_e (= \int I_e dt)$  is placed on a condenser of electrical capacitance  $C_e$ , the potential difference developed is given by

$$V_e = C_e^{-1} Q_e. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Similarly, if a quantity of heat  $Q_q$  is placed on a condenser of thermal capacitance  $C_q$ , the thermal potential difference developed is given by

$$V_q = \theta = C_q^{-1} Q_q. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The measure of  $C_q$ , according to this equation, is thus the product of the mass of the thermal condenser by its specific heat. Electrical capacitance may also be defined in terms of the potential energy stored in an electric field. The potential energy ( $P_e$ ) is given by

$$P_e = \frac{1}{2} C_e^{-1} Q_e^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

A thermal field also stores potential energy. If we write the expression analogous to equation (6), we get

$$P_q = \frac{1}{2} C_q^{-1} Q_q^2 = \frac{1}{2} \theta Q_q. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In this expression  $P_q$  has not the physical dimensions of energy but of energy multiplied by temperature. Equation (7) is thus false.

Similarly, an electric current flowing in an inductance sets up a magnetic field which stores kinetic energy in amount

$$K_e = \frac{1}{2} L_e I_e^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where  $K_e$  is the kinetic energy and  $L_e$  the (electric) inductance. Heat



flow under natural convection sets up a hydrodynamical field in which kinetic energy is stored and if, in analogy with equation (8), we write

$$K_q = \frac{1}{2} L_q I_q^2 \quad \dots \quad (9)$$

and define  $L_q$  by

$$\theta = L_q \frac{dI_q}{dt}, \quad \dots \quad (10)$$

we again find that  $K_q$  has the dimensions not of energy but of energy multiplied by temperature. It would appear, therefore, that one of the quantities, thermal current or thermal p.d., has not been properly defined.

Let us go back to consideration of flow in a simple resistance. The power developed when an electric current  $I_e$  flows against an electric p.d.  $V_e$  developed in a resistance  $R_e$  is given by

$$W_e = V_e I_e = R_e I_e^2 = V_e^2 / R_e \quad \dots \quad (11)$$

In the analogous heat-flow problem we might expect to write

$$W_q = \theta q a = R_q q^2 a^2 = \theta^2 / R_q \quad \dots \quad (12)$$

which is clearly incorrect because  $W_q$  here has the dimensions of power multiplied by temperature. If we wish to retain  $\theta$ , the temperature difference, as the thermal analogue of electrical p.d., then the thermal current density cannot be the quantity  $q$ , the heat crossing unit area in unit time, but must instead be a quantity of dimensions  $q/\theta$ . Let us write

$$I_q = q/\theta \quad \dots \quad (13)$$

Such a convention renders equations (7), (9) and (12) dimensionally correct, but instead of equation (3) we must now write

$$R_q = \frac{\theta}{q a / \theta} = \frac{\theta^2}{q a} \quad \dots \quad (14)$$

If the thermal resistance of a body is to be constant it then follows that  $q$ , the emittance, is proportional to the square of the temperature difference. There is no regular class of bodies known which follows such a law. Thus if equation (13) holds as a valid definition of thermal current, there is no thermal analogue of Ohm's law. The same definition of thermal current also leads to the conclusion that all thermal capacitances change rapidly with thermal potential differences. We therefore look for an alternative way of defining thermal current which shall be dimensionally equivalent to equation (13).

### *The Approach through Thermodynamics.*

When an electric current flows against an electric potential difference energy in an available form is converted into energy in an unavailable form, or work is converted into heat. Similarly, when heat flows down a temperature gradient, the associated energy becomes less available, or again, work is converted into heat.



The thermal current  $I_q$  in the variables given above becomes

$$I_q = \frac{qa}{T}. \quad (21)$$

The power developed when this current flows through the resistance  $R_q$  is thus

$$W = I_q^2 R_q = \frac{qa^2}{T}, \quad (22)$$

which is the correct expression for the rate of doing *work* by the flow of *heat*.

The specific thermal resistance of copper at 20° C. (293.1° K.) becomes, in terms of equation (20),

$$\sigma_q(\text{Cu}) = 75.6 \text{ thermal ohm/cm.},$$

which figure was obtained from that given by White (1938) by multiplying by 293.1. The corresponding figure for the specific thermal resistance of iron at 20° C. is

$$\sigma_q(\text{Fe}) = 469 \text{ thermal ohm/cm.}$$

In terms of the proposed new definition of thermal resistivity the Wiedemann and Franz law takes a particularly simple form ; namely, for any metal specimen the ratio of the electric resistance in ohms to the thermal resistance in thermal ohms is constant and independent of the size and shape of the body, its temperature or composition.

#### *The Thermal Farad.*

The unit of thermal capacitance may be defined in the following words :

A body has a capacitance of 1 thermal farad if an amount of entropy of 1 joule per degree C. added to that body raises the temperature by 1° C.

Thus, if a body of volume  $v$ , density  $\rho$ , and specific heat  $c_p$  (at constant pressure) has a temperature  $T$ , then the capacitance in thermal farads is (at constant pressure)

$$C_q = \frac{v\rho c_p}{T} \text{ (thermal farads)}. \quad (23)$$

For 1 c.c. of copper at 20° C. this capacitance amounts to

$$11.6 \text{ thermal millifarads,}$$

for 1 c.c. of iron at 20° C. to

$$12.7 \text{ thermal millifarads,}$$

and for 1 c.c. of water at 20° C. to

$$14.3 \text{ thermal millifarads.}$$

Thus thermal capacitances of bodies of ordinary dimensions are expressed by very much larger quantities than the corresponding electrical capacitances and hence, of course, thermal transients have much longer time constants than electrical transients. For non-inductive (thermal)

circuits the time constant is proportional to the product of the resistance (in thermal ohms) by the capacitance (in thermal farads). Since the first quantity involves the temperature of the body concerned in the numerator and the second in the denominator, the product is independent of the temperature, so that the same time constant for a non-inductive thermal circuit is obtained whether we use the definitions as given in equations (3) and (5) or the thermodynamically more exact definitions given in equations (19) and (23).

### *The Thermal Henry.*

If a heated body is placed in a fluid it sets up a system of convection currents in which heat is dissipated by means of fluid movement. Let the system of convection currents possess kinetic energy of amount  $K_q$  (in joules). The thermal inductance (Bosworth, 1946) may be defined by the expression

$$L_q = \frac{2K_q}{I_q^2} = \frac{2K_q T^2}{q^2 a^2} \text{ (thermal henrys).} \quad \dots \quad (24)$$

Accordingly, a figure may be given to  $L_q$  as soon as a quantitative measure has been given to the total kinetic energy stored in the circulating fluid system set up under natural convection.

### *Summary.*

The analogy between heat flow and electric current can only be upheld if the property analogous to electric charge is to be taken not as quantity of heat but as quantity of entropy. We then define the unit of thermal resistance in thermal ohms as the resistance of that body for which a temperature difference of  $1^\circ \text{C}$ . produces an entropy flow of 1 watt per degree C. With such a definition the rate of doing work by the flow of thermal current through a thermal resistance is

$$I_q^2 R_q,$$

completely analogous to the expression for the rate of doing work in an electric circuit.

A body may be defined as having a capacitance of 1 thermal farad if entropy in amount of 1 joule per degree C. raises the temperature by  $1^\circ \text{C}$ .

The system of convection currents surrounding a hot body in a fluid stores kinetic energy. A hot body immersed in a given fluid may, therefore, be described as having a thermal inductance of 1 thermal henry if an entropy flow of 1 watt per degree C. is associated with hydrodynamical kinetic energy of 1 joule.

### *References.*

- Avrami, M., and Paschkis, V., *Trans. Amer. Inst. Chem. Engrs.* xxxviii, pp. 631-52 (1942).  
 Harper, D. R., *Jour. Wash. Acad. Sci.* xviii, 17, pp. 459-71 (1928).  
 Paschkis, V., and Baker, H. D., *Trans. Amer. Soc. Mech. Engrs.* lxiv, pp. 105-112 (1942).  
 Paschkis, V., and Heisler, M. P., *J. of App. Phys.* xvii, pp. 246-54 (1946).  
 White, H. E., *Bull. Amer. Ceramic Soc.* xvii, pp. 17-20 (1938).  
 Bosworth, R. C. L., 'Nature,' clviii, 302 (1946).



LXXXIV. *A Note on Fusion and the Hole-Theory of Liquids.*

By A. G. CHOWDRI and F. C. AULUCK (University of Delhi)\*.

[Received June 25, 1945.]

DURING recent years the theory of the liquid state has attracted considerable attention, and many working models have been suggested to make the liquid state amenable to statistical treatment. In a number of papers Fürth<sup>(1)</sup> has developed the hole-model which idealises a liquid as a continuum permeated by holes and attributes the properties of the liquid to their presence.

Fürth's treatment is throughout classical. Recently we have constructed the Schrödinger equation for a hole and determined the eigenvalues of its energy<sup>(2)</sup>. In this paper we shall give some applications of the above results to the phenomenon of fusion. For the sake of completeness we shall first summarise the relevant results.

Consider a spherical hole of radius  $r$  in a material continuum of density  $\rho$ . We shall identify  $\rho$  with the density of the liquid. The formation of a hole requires an energy  $E_q$ ,

$$E_q = \frac{4\pi}{3}(p-p_0)r^3 + 4\pi r^2\sigma,$$

where  $\sigma$  is the surface tension of the liquid,  $p$  the external pressure, and  $p_0$  the saturation vapour pressure. The first term represents the work done against the pressure and the second that against the surface tension. A hole possesses four degrees of freedom, three of translation and one corresponding to a change in its radius. The kinetic energy of a hole can be written

$$\frac{p_x^2 + p_y^2 + p_z^2}{2m_1} + \frac{p_r^2}{2m_2},$$

where

$$m_1 = \frac{2m}{3}r^3\rho, \quad m_2 = 4\pi r^3\rho,$$

and the Hamiltonian of the problem as

$$H = \frac{p_r^2}{2m_2} + \frac{p_x^2 + p_y^2 + p_z^2}{2m_1} + 4\pi r^2\sigma + \frac{4\pi}{3}(p-p_0)r^3,$$

where  $m_1$  is the apparent mass of the hole for translation and  $m_2$  for radial motion. Neglecting the translatory part of the motion, the contribution of which to energy can be made to tend to zero by making the linear dimensions of the enclosure tend to infinity, we get the radial wave equation†

$$\frac{1}{r^6} \frac{d}{dr} \left( r^3 \frac{d\psi}{dr} \right) + \frac{8\pi\rho}{\hbar^2} \left\{ E - 4\pi r^2\sigma - \frac{4\pi}{3}(p-p_0)r^3 \right\} \psi = 0,$$

where  $E$  is an eigenvalue of  $H$ . We solve this equation by B.W.K.

\* Communicated by Professor D. S. Kothari.

† The contribution of the term  $\frac{4\pi}{3}(p-p_0)r^3$  is generally negligible.

Substance.	Triple point. $T_m$ °K.	Density $\rho$ (gm./cm. <sup>3</sup> ).	Surface tension $\sigma$ at $T_m$ °K (dynes/cm.).	$(\Delta V)_m$ volume (c.c.).	$\frac{\Theta}{T_m}$ .	$\frac{N_m}{N}$ .	$\frac{Q_m/RT_m}{\Theta \cdot \frac{N_m}{T_m}}$ .	$\frac{Q_m/RT_m}{(observed)}$ .
K	335.0	0.83	411.0	1.07	5.55	1.21	0.12	0.8
Na	371.0	0.93	294.0	0.51	3.82	0.525	0.42	0.85
Hg	230.0	14.2	500.0	0.52	4.13	2.16	0.13	1.2
Zn	693.0	6.7	770.0	0.69	2.31	2.50	0.21	1.2
Sn	500.0	6.98	518.0	0.46	2.39	1.44	0.48	1.65
Cd	594.0	8.0	600.0	0.65	2.15	2.29	0.24	1.2
Pb	600.0	10.3	440.0	0.66	1.58	2.26	0.29	1.05
Tl	561.0	11.0	400.0	0.55	1.60	1.86	0.23	0.7
P <sub>4</sub>	317.0	1.745	52.0	2.4	1.08	1.54	0.60	1.00
S <sub>8</sub>	388.0	1.81	59.0	7.2	0.96	4.96	0.66	3.12
Se <sub>2</sub>	473.0	4.50	92.5	2.82	0.84	3.48	0.96	2.80
Br <sub>2</sub>	266.0	3.1	44.0	8.16	0.97	6.24	0.79	4.80
N <sub>2</sub>	63.0	0.87	11.8	2.02	2.30	0.51	1.16	1.36
O <sub>2</sub>	54.3	1.283	22.0	1.64	3.73	0.64	0.40	0.96
H <sub>2</sub>	14.0	0.081	3.0	2.82	7.68	0.14	0.91	1.00
A <sub>1</sub>	83.8	1.4	13.5	3.05	1.66	1.0	0.96	1.60
Ne <sub>1</sub>	24.6	1.2	5.6	2.19	3.16	0.46	1.10	1.60
Ccl <sub>4</sub>	250.4	1.595	32.0	3.97	0.99	1.99	0.58	1.15
CS <sub>2</sub>	157.3	1.26	45.0	3.27	2.16	1.72	1.03	2.1
C <sub>6</sub> H <sub>6</sub>	278.4	0.878	30.9	10.28	1.04	3.93	0.60	4.2
C <sub>6</sub> H <sub>12</sub>	279.6	0.779	28.0	5.27	0.99	1.84	0.68	1.1
C <sub>6</sub> H <sub>5</sub> NH <sub>2</sub>	266.6	1.022	46.0	7.95	1.37	3.85	0.80	3.6
C <sub>6</sub> H <sub>5</sub> NO <sub>2</sub>	278.6	1.207	44.0	10.02	1.22	5.11	0.80	5.0
C <sub>6</sub> H <sub>4</sub> CH <sub>3</sub> NH <sub>2</sub>	316.6	1.046	37.0	15.1	0.98	6.73	0.80	5.3
C <sub>10</sub> H <sub>8</sub>	353.0	1.145	32.3	18.67	0.78	8.16	1.04	6.6
(C <sub>6</sub> H <sub>5</sub> ) <sub>2</sub> NH	327.0	1.159	38.0	16.2	0.94	7.63	0.89	6.4
(C <sub>6</sub> H <sub>5</sub> CH <sub>2</sub> ) <sub>2</sub>	325.2	0.942	32.0	16.8	0.89	6.72	1.42	8.5

method. We give here the final result of the calculations which give for the ground state

$$E_0 = 8.6 \frac{\hbar^{4/7} \sigma^{5/7}}{\rho^{2/7}},$$

and for the first excited state

$$E_1 = 14.1 \frac{\hbar^{4/7} \sigma^{5/7}}{\rho^{2/7}}.$$

Thus the "classical" radius of a hole-in-the-ground state is given by

$$E_0 = 4\pi\sigma r^2.$$

It is not unreasonable to assume that the whole of this energy is used for creating a hole. If  $\Theta$  denotes the temperature of melting, or more precisely the triple point where the pressure of the liquid is equal to the external pressure ( $p=p_0$ ), and  $k$  the Boltzmann constant,

$$\Theta = \frac{E_0}{k} = \frac{8.6}{k} \left( \frac{\hbar^4 \sigma^5}{\rho^2} \right)^{1/7}.$$

The column 6 in the table gives the values of the ratio  $\Theta/T_m$ , where  $T_m$  denotes the observed melting point. It may be observed that the ratio is not far from unity as required by the theory.

If some assumption be made regarding the number of holes created on melting, then we can obtain a relation between the heat of fusion  $Q_m$  and  $E_0$ . Let  $(\Delta V)_m$  represent the observed increase in volume (per gram-mole) on melting and  $v_0$  the average volume of a hole, then the number of holes per gram-hole is given by  $N_m = (\Delta V)_m / v_0$ ,

where  $v_0 = \frac{4\pi}{3} \left( \frac{E_0}{4\pi\sigma} \right)^{3/2}$ , and thus the heat of fusion  $Q_m$  is given by

$$\begin{aligned} Q_m &= E_0 N_m \\ &= E_0 \frac{(\Delta V)_m}{v_0}. \end{aligned}$$

Substituting the values of  $E_0$  and  $v_0$  we get

$$Q_m = 18.5 \sigma^{8/7} \rho^{1/7} (\Delta V)_m \text{ joules.}$$

The latent heat per hole is therefore given by

$$\frac{Q_m}{N_m} = E_0 = k\Theta,$$

or

$$Q_m / RT_m = k\Theta N_m / RT_m = \frac{N_m}{N} \cdot \frac{\Theta}{T_m},$$

and hence

$$\frac{Q_m}{RT_m} \cdot \frac{\Theta}{T_m} \cdot \frac{N_m}{N} = 1.$$

The values of

$$\frac{N_m}{N} \quad \text{and} \quad \frac{Q_m}{RT_m} \cdot \frac{\Theta}{T_m} \cdot \frac{N_m}{N}$$

are given in the columns 7 and 9 respectively of the table. According to theory the values in the last column should be unity. The table shows that the agreement between theory and observation, considering the crude nature of the theory, is not unsatisfactory.

#### References.

- (1) Fürth, R., Proc. Camb. Phil. Soc. xxxvii. p. 252 (1941).
- (2) F. C. Auluck and D. S. Kothari, Proc. Camb. Phil. Soc. xli. p. 180 (1945).

LXXXV. *Notices respecting New Books.*

*The Methods of Plane Projective Geometry Based on the Use of General Homogeneous Co-ordinates.* By E. A. MAXWELL. [Pp. xix+230.] (Cambridge University Press, 1946.) Price 12s. 6d.

THIS very well-written book emphasizes methods rather than results, and some of the algebraic technique which has been employed in Cambridge for many years is here published for the first time.

The first three chapters are given to homogeneous co-ordinates, one-one algebraic correspondence, cross-ratio and harmonic ranges. The next seven chapters develop the properties of conics and pencils of conics, including the subject of correspondence on a conic. The ideas of length and angle are introduced in Chapter XI., where the relationship between Euclidean and Projective Geometry is explained. Projective results are interpreted metrically and applied to many problems in Chapter XII. A large number of examples, taken mostly from examination papers, are included at the end of most chapters.

The book is intended for first-year University classes and should prove valuable to all students who wish to learn the best methods.

L. S. GODDARD.

*A First Course in Mathematical Statistics.* By C. E. WEATHERBURN. [Pp. xv+271.] (Cambridge University Press, 1946.) Price 15s. 0d.

IT is a pleasure to find yet another book written by Professor Weatherburn. His style is too well-known to need comment, and in the present book all the good features of the author's previous books, on the Vector Calculus and on Differential Geometry, are maintained. The applications of mathematical statistics are now very wide, ranging as they do from such subjects as Agriculture and Biology to Economics and Psychology, and Professor Weatherburn's book should appeal to those who, while not being mathematical specialists, find it necessary to know something of the mathematical principles underlying the interpretation of statistical data.

The book arose out of a course of lectures given to mixed classes of students, who were assumed to have an ordinary knowledge of mathematics, including the integral calculus. The first five chapters deal with Frequency Distributions, Probability Distributions and some Standard Distributions (Binomial, Poisson, Normal). The theory of Regression and Correlation is also developed. Chapter VI. deals with the Theory of Simple Sampling, and Chapter VII. with the Standard Errors of Statistics. The properties of beta and gamma variates and their distributions are treated in Chapter VIII. and used in the next two chapters in proving the sampling distributions of the statistics, which form the basis of the common tests of significance. Chapter XI. deals with the Analysis of Variance and Covariance, and Chapter XII. with Multivariate Distributions. A large number of examples illustrating the theory appear in the text and a list of exercises is given at the end of each chapter.

The book is an excellent textbook for classes of students studying statistics for the first time.

L. S. GODDARD.

---

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]



LXXXVI. *On the Bending of an Initially Straight Beam Under Arbitrary Loading.*

By A. W. GILLIES (Northampton Polytechnic) \*.

[Received June 21, 1946.]

*Abstract.*

The general solution for the bending of an initially straight beam under arbitrary loading is obtained in a form which it is suggested is convenient in application and easy to remember. A number of examples are worked out to illustrate the application of the general solution to particular problems.

The method may be regarded as an extension of Macaulay's method.

---

*Introduction.*

THE bending of an initially straight beam under various conditions of loading has been treated so often that some explanation is required to justify yet another paper on this subject.

The available texts seem either to deal with each variation of loading conditions as a separate problem using different methods for each set of conditions, or on the other hand the treatment is based on the principle of superposition, the more complicated cases being obtained by superposition of simple cases, so that the results of many simple cases have to be remembered.

There seems, therefore, to be room for a uniform analytical treatment of the general case, if the result can be expressed in a form which is easy to remember and convenient in use.

The present paper is an attempt to fulfil this need.

*Sign Convention and Symbols.*

It is impossible to discuss such a problem as this in general form without some definite convention as to signs and symbols. The following conventions are therefore adopted. The centre line of the beam is assumed to lie initially along the  $x$ -axis. The  $y$ -axis will be chosen (fig. 1) to form a right-handed system so that in normal gravity loading the rate of loading will be negative. The positive sense of rotation is that which takes the positive sense of the  $x$ -axis into the positive sense of the  $y$ -axis by a rotation of one right angle, *i. e.* is anticlockwise. An anticlockwise moment will therefore be considered positive and a transverse force in the positive sense of the  $y$ -axis will be considered positive.

The stresses across any section of the beam are equivalent to a force and a couple acting on the material to the left of the section *due to the*

---

\* Communicated by the Author.

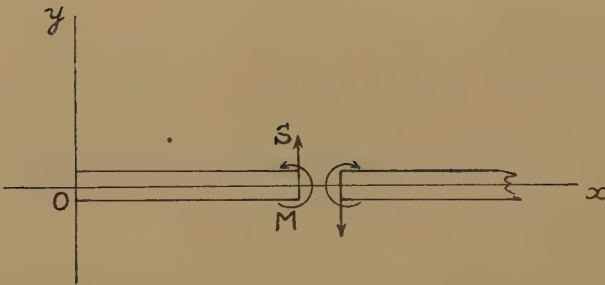
material to the right, and an opposite force and couple acting on the material to the right of the section *due to* the material to the left. This state of stress will be specified for a section perpendicular to the  $x$ -axis by the force (stress resultant or shearing force  $S$ ) and couple (stress couple or bending moment  $M$ ) on the material to that side of the section for which the outward drawn normal is in the positive sense of the  $x$ -axis.

This gives the scheme shown in fig. 1.

It can be shown that with these conventions the following relations are valid :—

$$\begin{aligned} M &= EI \frac{d^2y}{dx^2}, \\ -S &= \frac{dM}{dx} = EI \frac{d^3y}{dx^3}, \\ w &= -\frac{dS}{dx} = \frac{d^2M}{dx^2} = EI \frac{d^4y}{dx^4}. \end{aligned}$$

Fig. 1.



If a concentrated load  $F$  is applied at  $x=a$ , we have a discontinuity in shearing force given by  $S_{a-} = S_{a+} + F$ , where  $S_{a-}$  and  $S_{a+}$  denote the values of the shearing force on the sides of the load for which  $x$  is just less than  $a$ , and just greater than  $a$ , respectively.

Similarly, if a concentrated couple of moment  $G$  is applied at  $x=b$ , there is a discontinuity of bending moment of amount given by  $M_{b-} = M_{b+} + G$ .

#### Single Concentrated Load.

Consider a portion of beam having a single concentrated load  $F$  at  $x=a$  and let the subscript 0 indicate values at the origin ; then consider the equilibrium of the portion of the beam between 0 and  $x$ ,  $x$  and  $a$  both being supposed positive.

Resolving the forces on the beam between 0 and  $x$  parallel to the  $y$ -axis, we have

$$S - S_0 + F\{x-a\}^0 = 0. \quad \dots \dots \dots (1)$$

In this and subsequent equations the bracket  $\{ \}$  is to be interpreted as in Macaulay's method, namely, that  $\{z\}^n = z^n$  if  $z$  is positive, but  $\{z\}^n = 0$  if  $z$  is negative. Thus, if  $x > a$ ,  $\{x-a\}^0 = (x-a)^0 = 1$ , while if  $x < a$ ,

$\{x-a\}^0=0$ , the force  $F$  not then acting on the portion of the beam under consideration.

The integral between 0 and  $X$  of terms of the form  $\{x-a\}^n$  will be required. We shall have

$$\int_0^X \{x-a\}^n dx = 0$$

if  $X < a$ , since the integrand is then zero for all values of  $x$  in the interval of integration.

On the other hand, if  $X > a$ , we have

$$\begin{aligned} \int_X^0 \{x-a\}^n dx &= \int_0^a \{x-a\}^n dx + \int_a^X \{x-a\}^n dx \\ &= 0 + \int_a^X (x-a)^n dx \\ &= \left[ \frac{(x-a)^{n+1}}{n+1} \right]_a^X \\ &= \frac{(X-a)^{n+1}}{n+1} - 0 \\ &= \frac{(X-a)^{n+1}}{n+1} \end{aligned}$$

Thus, both cases are included in the single formula

$$\int_0^X \{x-a\}^n dx = \frac{\{X-a\}^{n+1}}{n+1}.$$

Returning now to equation (1), writing  $-\frac{dM}{dx}$  for  $S$  and integrating between 0 and  $X$ , we have

$$-\int_0^X \frac{dM}{dx} dx - \int_0^X S_0 dx + \int_0^X F \{x-a\}^0 dx = 0,$$

$$\text{or} \quad -(M_X - M_0) - S_0 X + F \frac{\{X-a\}^1}{1} = 0.$$

Since  $X$  may have any value, we may now replace it by  $x$ , giving

$$-M + M_0 - S_0 x + F \frac{\{x-a\}^1}{1} = 0. \quad \dots \quad (2)$$

Writing  $M = EI \frac{d\theta}{dx}$ , where  $\theta \doteq \tan \theta = \frac{dy}{dx}$ , integrating between 0 and  $X$  again, and replacing  $X$  by  $x$  after the integration is completed gives, in exactly the same way,

$$-EI\theta + EI\theta_0 + M_0 x - S_0 \frac{x^2}{2!} + F \frac{\{x-a\}^2}{2!} = 0, \quad \dots \quad (3)$$

and a final integration, after replacing  $\theta$  by  $\frac{dy}{dx}$ , gives

$$-EIy + EIy_0 + EI\theta_0 x + M_0 \frac{x^2}{2} - S_0 \frac{x^3}{3!} + F \frac{\{x-a\}^3}{3!} = 0. \quad \dots \quad (4)$$





A load  $w\delta\xi$  at  $x=\xi$  will thus give a term  $w\delta\xi \frac{\{x-\xi\}^n}{n!}$ .

The total contribution from the distributed load will therefore be

$$\Sigma w \cdot \frac{\{x-\xi\}^n}{n!} \delta\xi,$$

or in the limit

$$\int_0^x w \cdot \frac{\{x-\xi\}^n}{n!} \cdot d\xi.$$

Since  $\{x-\xi\}^n$  is zero when  $\xi > x$ , we need only sum from 0 to  $x$ , giving

$$\int_0^x w \cdot \frac{(x-\xi)^n}{n!} \cdot d\xi,$$

and since the limits ensure that  $(x-\xi)^0$  is positive throughout the interval of integration, the brackets  $\{\}$  may be dispensed with and ordinary brackets used. The previous form is, however, sometimes useful.

### General Case of Arbitrary Loading.

Consider now a beam in which all forms of loading are present simultaneously. Suppose there are concentrated loads  $F_r$  at

$$x=a_r(r=1, 2, \dots n),$$

concentrated couples  $G_s$  at  $x=b_s(s=1, 2, \dots m)$ , and a distributed load of intensity  $w$ . Terms of all the types considered will be present in the equations and, assembling the previous results, we shall obtain for this case the equations

$$-S = -S_0 + \sum_{r=1}^n R_r \{x-a_r\}^0 + \int_0^x w dx, \quad \dots \dots \dots (1b)$$

$$M = M_0 - S_0 x + \sum_{r=1}^n F_r \frac{\{x-a_r\}^1}{1!} - \sum_{s=1}^m G_s \{x-b_s\}^0 + \int_0^x w(x-\xi) d\xi, \quad \dots \dots (2b)$$

$$EI\theta = EI\theta_0 + M_0 x - S_0 \frac{x^2}{2!} + \sum_{r=1}^n F_r \frac{\{x-a_r\}^2}{2!} - \sum_{s=1}^m G_s \frac{\{x-b_s\}^1}{1!} + \int_0^x w \frac{(x-\xi)^2}{2!} d\xi, \\ \dots \dots \dots (3b)$$

$$EIy = EIy_0 + EI\theta_0 x + M_0 \frac{x^2}{2!} - S_0 \frac{x^3}{3!} + \sum_{r=1}^n F_r \frac{\{x-a_r\}^3}{3!} \\ - \sum_{s=1}^m G_s \frac{\{x-b_s\}^2}{2!} + \int_0^x w \frac{(x-\xi)^3}{3!} d\xi. \quad \dots \dots (4b)$$

These four equations give the general solution of the problem in terms of the four constants  $S_0$ ,  $M_0$ ,  $\theta_0$  and  $y_0$ .

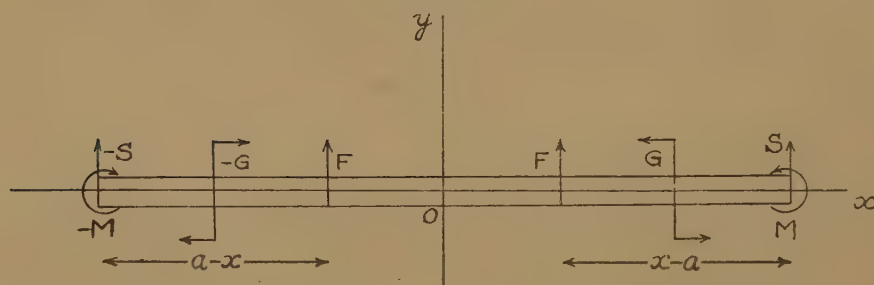
Several features of this solution may be noted which should make it easier to remember. Firstly, the last equation only need be remembered since the previous ones can easily be recovered by differentiation with

respect to  $x$ . In this connection it will be noted that the last term can be differentiated with respect to  $x$  under the sign of integration, *i. e.* the integrand of the  $\xi$  integral may be differentiated with respect to  $x$ . While this is generally permissible when the limits are constants, it is not usually legitimate when the limits are variable, as in this case, when the upper limit is  $x$ . The correctness of the result depends on the fact that the integrand is zero at the upper limit.

The terms arising from the forces  $S_0$ ,  $F_r$ ,  $wd\xi$  are all of the form Force  $X$  (length)<sup>3</sup>/3! while the terms arising from the moments  $M_0$  and  $G$  are of the form Moment  $X$  (length)<sup>2</sup>/2! The  $S_0$  and  $M_0$  terms are opposite in sign to the other terms in these respective groups, because they are opposite to external force and couple which will have to be applied at the origin if the beam terminates there. Again, if the way in which  $S$ ,  $M$ ,  $\theta$  and  $y$  are connected by successive integrations is kept in mind the powers of  $x$  and factorials associated with  $S_0$ ,  $M_0$ ,  $\theta_0$  and  $y_0$  should be easily recollected.

Finally, in applying the solution to specific problems, the four constants  $S_0$ ,  $M_0$ ,  $\theta_0$ ,  $y_0$  have to be determined to satisfy particular conditions of

Fig. 2.



support, and it may be noted that in nearly every case two of them can be written down at once. If the beam is simply supported at 0 we have  $y_0=M_0=0$ . If 0 is a built-in end, then  $y_0=\theta_0=0$ , while at a free overhanging end  $S_0=M_0=0$ .

It should be emphasized that the solution obtained applies to the portion of the beam to the right of the origin, *i. e.* for  $x$  positive, and the summations only apply to loads to the right of the origin. The equations require slight modification if the origin is not at the left-hand end of the beam, and negative values of  $x$  are being considered.

Probably in most cases the left-hand end of the beam will be the most convenient point for origin, but cases certainly do arise when an intermediate point is preferable, so the modification should be noted.

The equations may be re-written from the start and the solution derived in a similar manner. It will be better, however, from the point of view of remembering the result, to deduce it from considerations of symmetry. Fig. 2 shows that symmetry requires that  $\{x-a_r\}$  and  $\{x-b_s\}$  should be replaced by  $\{a_r-x\}$  and  $\{b_s-x\}$  respectively, which will give non-zero

values according to the convention when  $F_r$  and  $G_s$  act on the portion of the beam between 0 and a negative value of  $x$ , while the sign  $G_s^i$  must be reversed so that for  $x$  negative the equation becomes

$$EIy = EIy_0 + EI\theta_0 x + M_0 \frac{x^2}{2} - S_0 \frac{x^3}{6} + \Sigma F_r \frac{\{a_r - x\}^3}{6} \\ + \Sigma G_s \frac{\{b_s - x\}^2}{2} + \int_0^x w \frac{(x - \xi)^3}{6} d\xi. \quad (4c)$$

It will be noted that symmetry does not require the reversal of sign of the  $\theta_0$  and  $S_0$  terms since this is already accomplished by the odd powers of  $x$  which multiply these quantities, while in the distributed load integral  $d\xi$  is negative due to the negative upper limit. The last equation is thus only altered in the manner described in the terms involving concentrated loads and moments. The first three equations are obtained by differentiation from (4c).

The terms which involve the distributed load have been deduced from those for a single concentrated load. They may also be obtained by carrying out the actual integrations by which each successive equation is derived from the previous one. Thus in (1b) we have the last term

$\int_0^x w dx$ , and (2b) is derived by integrating (1b) between 0 and  $x$ , which

would give for the last term  $\int_0^x dx \int_0^x w d\xi$ , which must therefore equal

$\int_0^x w(x - \xi) d\xi$ . Similarly, the next integration to obtain (3b) would give

$\int_0^x dx \int_0^x dx \int_0^x w d\xi$ , which must equal  $\int_0^x w \frac{(x - \xi)^2}{2!} d\xi$ , and the final inte-

gration which gives (4b) would give for the last term

$$\int_0^x dx \int_0^x dx \int_0^x dx \int_0^x w d\xi, \text{ which must equal } \int_0^x w \frac{(x - \xi)^3}{3!} d\xi.$$

The equality of these various pairs is demonstrated in the Appendix. It may be remarked here that it is sometimes more convenient to use the single integral and in other cases it may be easier to carry out the four successive integrations. This will be seen in the examples which are given illustrating the application of the general formula.

It has already been remarked that in almost all cases two of the integration constants can be written down at once, so that usually only two have to be calculated. The application of the result will now be considered in a selection of problems, some very simple and as easily worked from first principles and others more involved. It is hoped that these examples will provide evidence of the convenience of the general solution which has been obtained.

The special features to which attention may be drawn are: (i) the use of the curly brackets, commonly known as Macaulay's Method; (ii) the advantage of making the integrations definite between limits

0 and  $x$ ; (iii) the expression of the term arising from the distributed load as a single integral.

1. *Single Concentrated Load on a Beam with Built-in Ends* (fig. 3).

Here  $y_0 = \theta_0 = 0$  and

$$EIy = M_0 \frac{x^2}{2} - S_0 \frac{x^3}{6} - F \frac{\{x-a\}^3}{6}.$$

$M_0$  and  $S_0$  are determined from the condition  $y = \theta = 0$  at  $x = l$ .

Substituting  $x = l$ ,

$$0 = M_0 \frac{l^2}{2} - S_0 \frac{l^3}{6} - F \frac{b^3}{6}.$$

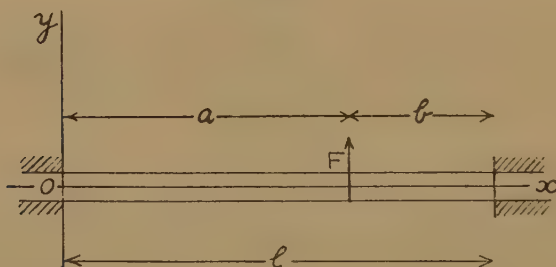
Differentiating and putting  $x = l$ ,

$$0 = M_0 l - S_0 \frac{l^2}{2} - F \frac{b^2}{2}.$$

Solving for  $M_0$  and  $S_0$  gives

$$S_0 = - \frac{Fb^2(3l-2b)}{l^3} \quad M_0 = - \frac{Fb^2(l-b)}{l^2}.$$

Fig. 3.



Substituting these values in the first equation gives

$$EIy = - \frac{Fb^2a}{l^2} \cdot \frac{x^2}{2} + \frac{Fb^2(2l-2b)}{l^3} \cdot \frac{x^3}{6} - F \frac{\{x-a\}^3}{6}.$$

For the section OA the deflection is given by

$$EIy = - \frac{Fb^2a}{l^2} \cdot \frac{x^2}{2} + \frac{Fb^2(3l-2b)}{l^3} \cdot \frac{x^3}{6}.$$

It may be well to emphasize that  $M_0$  and  $S_0$  are the bending moment and shearing force at  $x=0$ , and since the force and couple acting on the beam due to the support are forces acting on the material to the right of the section, their values will be  $-S_0$  and  $-M_0$  respectively, as can easily be verified directly.

2. *Simply Supported Beam Bent by a Couple at a Point on its Length.*

Take origin at A, then

$$y_0 = M_0 = 0.$$

$$EIy = EI\theta_0 x - S_0 \frac{x^3}{6} - G \frac{\{x-a\}^2}{2}.$$



$y=0$  and  $M=0$  at  $x=l$  gives

$$0 = EI\theta_0 l - S_0 \frac{l^3}{6} - G \frac{b^2}{2}.$$

$$0 = -S_0 l - G.$$

$$\therefore S_0 = -\frac{G}{l} \text{ and } EI\theta_0 = \frac{G}{6l}(3b^2 - l^2).$$

Then

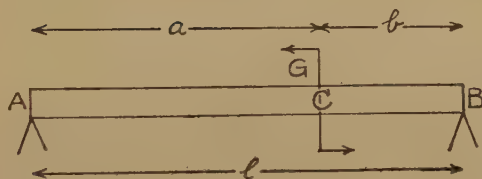
$$EIy = \frac{Gx}{6l}(3b^2 - l^2) + \frac{Gx^3}{6l} - G \frac{\{x-a\}^2}{2}.$$

Differentiating and putting  $x=l$  gives

$$\begin{aligned} EI\theta_B &= \frac{G}{6l}(3b^2 - l^2) + \frac{Gl}{2} - Gb \\ &= \frac{G}{6l}[3b^2 - l^2 + 3l^2 - 6lb] \\ &= \frac{G}{6l}[3a^2 - l^2], \end{aligned}$$

while  $\theta_A = \theta_0$ .

Fig. 4.



$$\text{Thus } \theta_A = \frac{G(3b^2 - l^2)}{6EI} \quad \theta_B = \frac{G(3a^2 - l^2)}{6EI}$$

and on AC

$$y = \frac{Gx}{6EI}(3b^2 - l^2 + x^2).$$

3. A simply-supported beam of length  $l$  carries a uniform distributed gravity load over the portion  $x=a$  to  $x=b$  (rate of loading  $-w$ ).

$$\text{Here we have } EIy = EI\theta_0 x - S_0 \frac{x^3}{3!} - \int_a^b w \frac{\{x-\xi\}^3}{3!} d\xi.$$

In the integral the curly bracket ensures that only elements  $d\xi$  between 0 and  $x$  are considered, while the limits  $a$  and  $b$  arise from the fact that the rate of loading  $w$  is zero outside this range. Carrying out the integration gives at once

$$EIy = EI\theta_0 x - S_0 \frac{x^3}{3!} - w \frac{\{x-a\}^4}{4!} + w \frac{\{x-b\}^4}{4!}.$$

The conditions  $y=0$  and  $M=0$  at  $x=l$  then give

$$S_0 = -\frac{w}{2l}(2l-a-b)(b-a)$$

$$\text{and } EI\theta_0 = -\frac{w}{24l}(2l-a-b)(b-a)[4l^2 - 2l(a+b) + a^2 + b^2],$$

so that finally

$$EIy = -\frac{w}{24l}(2l-a-b)(b-a)[x(4l^2-2la-2lb+a^2+b^2)-2x^3] \\ -\frac{w}{24}\{x-a\}^4 + \frac{w}{24}\{x-b\}^4.$$

On putting  $a=0$ ,  $b=l$ , this reduces to the solution for a uniformly loaded beam.

#### 4. Theorem of Three Moments.

Here we wish to relate the bending moments at three successive supports A, O, and B on a continuous beam with uniform loading  $w$ .

Take the origin at the middle support and consider OB,

$$EIy = EI\theta_0 x + M_0 \frac{x^2}{2} + S_0 x + w \frac{x^4}{24}.$$

Putting  $x=b$ ,

$$EIy_B = EI\theta_0 b + M_0 \frac{b^2}{2} - S_0 b + w \frac{b^4}{24}.$$

Differentiate twice and put  $x=b$ .

$$M_B = M_0 - S_0 b + w \frac{b^2}{2},$$

$$i. e. \quad M_B \frac{b^2}{6} = M_0 \frac{b^2}{6} - S_0 b + w \frac{b^4}{12}.$$

Subtract the previous equation to eliminate  $S_0$  gives

$$M_B \frac{b^2}{6} - EIy_B = -EI\theta_0 b - M_0 \frac{b^2}{3} + w \frac{b^4}{24}.$$

Since there are no concentrated loads between A and B except the reaction at O, the same general equation will apply to OA if  $S_0$  is replaced by  $S_0$ . If then we carry out the same steps for OA, the result will be obtained by replacing  $b$  by  $-a$ ,  $M_B$  by  $M_A$  and  $y_B$  by  $y_A$ . Thus we obtain

$$M_A \frac{a^2}{6} - EIy_A = +EI\theta_0 a - \frac{M_0 a^2}{3} + w \frac{a^4}{24}.$$

Dividing this equation by  $a$  and the previous equation by  $b$  and adding eliminates  $EI\theta_0$  and leads to

$$M_A \frac{a}{6} + M_B \frac{b}{6} - EI \left( \frac{y_A}{a} + \frac{y_B}{b} \right) = -M_0 \frac{(a+b)}{3} + w \frac{(a^3+b^3)}{24}$$

$$\text{or} \quad aM_A + 2(a+b)M_0 + bM_B = \frac{w}{4}(a^3+b^3) + 6EI \left( \frac{y_A}{a} + \frac{y_B}{b} \right).$$

This is the theorem of the three moments.

If the supports are all at the same level, it takes the simpler form on putting  $y_A = y_B = 0$ .

$$aM_A + 2(a+b)M_0 + bM_B = \frac{w}{4}(a^3+b^3).$$

5. A uniform beam of length  $2l$  is simply supported at its ends and carries a symmetrically distributed load of intensity

$$w \left( 1 - \frac{x^2}{l^2} \right)$$

at a distance  $x$  from the centre. Determine the maximum deflection.

Take the origin at the centre and consider the right-hand half. From symmetry  $S_0 = \theta_0 = 0$  and at  $x=l$  we have  $y=M=0$ .

Here it is convenient to evaluate the term arising from the distributed load by successive integration, so we start with the equation for  $S$ , which with  $S_0=0$  is

$$\begin{aligned} -S &= \int_0^x w \left( 1 - \frac{\xi^2}{l^2} \right) d\xi \\ &= wx - \frac{w x^3}{3l^2}. \end{aligned}$$

Integrating between 0 and  $x$  gives

$$M = M_0 + \frac{wx^2}{2} - \frac{wx^4}{12l^2}.$$

$x=l$  gives

$$0 = M_0 + \frac{wl^2}{2} - \frac{wl^2}{12}, \text{ i. e. } M_0 = -\frac{5wl^2}{12}.$$

Integrating again twice gives

$$EI\theta = -\frac{5wl^2x}{12} + \frac{wx^3}{6} - \frac{wx^5}{60l^2}, \text{ since } \theta_0 = 0,$$

and

$$EIy = EIy_0 - \frac{5wl^2x^2}{24} + \frac{wx^4}{24} - \frac{wx^6}{360l^2}.$$

$x=l$  gives

$$0 = EIy_0 + \left( -\frac{5}{24} + \frac{1}{24} - \frac{1}{360} \right) wl^4,$$

$$\text{i. e. } EIy_0 = \frac{61wl^4}{360}$$

and

$$EIy = \frac{61wl^4}{360} - \frac{5wl^2x^2}{24} + \frac{wx^4}{24} - \frac{wx^6}{360l^2}.$$

The deflection is clearly a maximum at  $x=0$  since  $\theta_0=0$  and  $M_0$  is negative, therefore we have

$$y_{\max} = \frac{61wl^4}{360EI}.$$

If the loading is gravitational  $w$  will be negative, and this value will be the numerically greatest deflection, and it will be downwards.

6. A uniform beam is supported at its extremities and at two other points dividing the beam into three equal spans, all four being at the same level. Prove that the pressures on the supports are as 4 : 11 : 11 : 4 (fig. 5).

Let the reactions be  $P$  and  $Q$ , as shown. Then, taking origin at end of beam,  $y_0 = M_0 = 0$  and  $S_0 = -P$ .

$$EIy = EI\theta_0 x + P \frac{x^3}{6} + Q \frac{\{x-l\}^3}{6} + Q \frac{\{x-2l\}^3}{6} - w \frac{x^4}{24}.$$

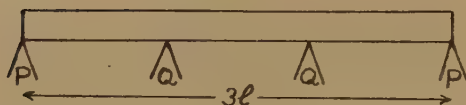
Substitute  $y=0$  for  $x=l, 2l, 3l$  in turn gives

$$0 = EI\theta_0 \cdot l + P \frac{l^3}{6} - w \cdot \frac{l^4}{24} \quad \dots \quad (1)$$

$$0 = EI\theta_0 \cdot 2l + P \frac{8l^3}{6} + Q \frac{l^3}{6} - w \cdot \frac{16l^4}{24} \quad \dots \quad (2)$$

$$0 = EI\theta_0 \cdot 3l + P \frac{27l^3}{6} + Q \cdot \frac{8l^3}{6} + Q \cdot \frac{l^3}{6} - w \cdot \frac{81l^4}{24} \quad \dots \quad (3)$$

Fig. 5.



From new equations (2)  $-2 \cdot (1)$  and (3)  $-(2)-(1)$  :—

$$0 = +Pl^3 + Q \frac{l^3}{6} - w \cdot \frac{14l^4}{24}.$$

$$0 = +P \cdot 3l^3 + Q \cdot \frac{8l^3}{6} - w \cdot \frac{64l^4}{24}.$$

Multiply the first of these by 32 and the second by 7 and subtract :—

$$0 = +P \cdot 11l^3 - Q \cdot 4l^3,$$

i. e. 
$$\frac{P}{Q} = + \frac{4}{11},$$

and ratio 4 : 11, as required.

7. A uniform heavy beam is supported at its extremities A and C and at its middle point B ; A and C are at the same level and B is such that the pressures on the three supports are equal. Prove that the depth of B below AC is  $\frac{7}{15}$  of the whole central deflection when the beam is supported only at A and C.

The reactions are each  $\frac{wl}{3}$ .

Taking origin at A,  $y_0 = M_0 = 0$  and  $S_0 = -\frac{wl}{3}$ .

$$\therefore EIy = EI\theta_0 x + \frac{wl}{3} \frac{x^3}{6} + \frac{wl}{3} \frac{\{x-\frac{1}{2}\}^3}{6} - \frac{wx^4}{24}.$$



$y=0$  when  $x=l$  gives

$$0 = EI\theta_0 l + wl^4 \left[ \frac{1}{18} + \frac{1}{18 \cdot 8} - \frac{1}{24} \right].$$

$$\therefore EI\theta_0 = -\frac{wl^3}{144} [8+1-6] = -\frac{wl^3}{48}.$$

$$\therefore EIy = -\frac{wl^3x}{48} + \frac{wlx^3}{18} + \frac{wl}{18} \left\{ x - \frac{l}{2} \right\}^3 - \frac{wx^4}{24}.$$

$x = \frac{l}{2}$  then gives

$$\begin{aligned} EIy_B &= -wl^4 \left[ \frac{1}{96} - \frac{1}{144} + 0 + \frac{1}{384} \right] \\ &= -wl^4 \cdot \frac{1}{1152} (12-8+3) \\ &= -\frac{7wl^4}{1152}. \end{aligned}$$

When the beam is supported at A and C only the central deflection is

$$-\frac{5wl^4}{384EI}$$

and ratio is  $\frac{7}{1152} \div \frac{5}{384} = \frac{7}{15}$  as required.

8. (i) A section AB of a beam is loaded and deflected in any manner. It is required to produce an additional rotation  $\Theta$  of the end B without any additional deflection at B or any additional deflection or rotation of the end A. Since the equations are linear we need only consider the additional loads and deflections. Taking origin at A, since there are no additional loads between A and B and  $y_0 = \theta_0 = 0$ , the deflection equation will be

$$EIy = M_0 \frac{x^2}{2!} - S_0 \frac{x^3}{3!},$$

$y=0$  when  $x=l$  gives  $S_0 l = 3M_0$ ,

$$EI\theta = M_0 x - S_0 \frac{x^2}{2!},$$

$\theta = \Theta$  when  $x=l$  gives  $EI\Theta = -\frac{M_0 l}{2}$ ,

$$M = M_0 - S_0 x.$$

If the external couple at B is  $G$ , then

$$G = M_l = -2M_0.$$

The external couple required at end A,  $G_T$ , say, is  $-M_0$ , i. e.  $G_T = \frac{1}{2}G$ , while, in terms of  $\Theta$ , we have

$$G = \frac{4EI\Theta}{l} = 4EK\Theta,$$

in which  $K = \frac{1}{l}$  is a constant of the beam, independent of the loading, known as the stiffness factor.

Thus, if a rotation  $\Theta$  is produced at one end of a beam without change of deflection, the other end remaining fixed in position and direction, the external couple required is proportional to  $EK\Theta$  and an external couple of half the magnitude, and in the same sense will be required at the fixed end.

(ii) If an additional rotation is produced at end B, while end A remains fixed in position but not in direction, we have  $y_0 = M_0 = 0$  and

$$EIy = EI\theta_0 x - S_0 \frac{x^3}{3!},$$

$$y = 0 \text{ when } x = l \text{ gives } S_0 = \frac{6EI}{l^2} \theta_0,$$

$$EI\theta = EI\theta_0 - S_0 \cdot \frac{x^2}{2!}$$

and  $\theta = \Theta$  when  $x = l$  gives  $\theta_0 = -\frac{1}{2}\Theta$ ,

$$M = -S_0 x,$$

i. e.  $M_l = -S_0 l - \frac{6EI}{l} \Theta.$

The external couple required at B is

$$G = M_l = -\frac{6EI\theta_0}{l} = \frac{3EI}{l} \Theta, \text{ i. e. } G = 3EK\Theta.$$

Thus the external moment required in this case is 3/4 of the moment required in the previous case.

(iii) If, instead of producing a rotation of the end B, a deflection  $\Delta$  is produced without rotation, the end A being fixed as in the first case, then the equations are

$$EIy = M_0 \frac{x^2}{2!} - S_0 \frac{x^3}{3!},$$

$$y = \Delta \text{ when } x = l \text{ gives } EI\Delta = M_0 \frac{l^2}{2} - S_0 \frac{l^3}{6},$$

whilst  $EI\theta = M_0 x - S_0 \frac{x^2}{2!}$

and  $\theta = 0$  when  $x = l$  gives  $M_0 = \frac{S_0 l}{2},$

so that  $EI\Delta = \frac{1}{6} M_0 l^2.$

Also  $M = M_0 - S_0 x,$

so that  $M_l = M_0 - S_0 l = -M_0$

and  $G = M_l = -\frac{6EI\Delta}{l^2}.$

and  $G_T = -M_0 = G.$

Thus, when one end of a beam is deflected through a distance  $\Delta$  without rotation, the other end remaining fixed in position and direction, equal external moments are required at the ends of values equal to

$$-\frac{6EI\Delta}{l^2}.$$

(iv) The three results just obtained are those on which is based the "Moment Distribution Method" of analysing rigid frames. The method consists in dividing a continuous beam into sections and imagining external moments applied at the ends of each section to prevent deflection or rotation of the end. No matter how complicated the case under consideration, it can always be divided in such a way that each section is a simple case of an encastré beam for which the solution is known. We shall then have the solution for the beam subject to the actual loading together with additional external moments at the boundaries of the sections. If AB, BC are two sections and  $M_{B-}$ ,  $M_{B+}$  are the bending moments on either side of B, the external moment required at B is then  $G_B = M_{B-} - M_{B+}$ .

We now remove this moment, *i. e.* apply an external moment  $-G_B$  at B in such a way as to produce a rotation  $\Theta$  at B. If the additional bending moments are  $M'_{B-}$  and  $M'_{B+}$  and no change in position or direction is produced at the other ends of the sections, then we have

$$\left. \begin{aligned} -G_B &= M'_{B-} - M'_{B+}, \\ M'_{B-} &= 4E_1K_1\Theta \\ -M'_{B+} &= 4E_2K_2\Theta \end{aligned} \right\} \text{by the result of case (i) above,}$$

$$\text{adding, } M'_{B-} - M'_{B+} = 4(E_1K_1 + E_2K_2)\Theta = -G_B.$$

$$\text{i. e. } \Theta = \frac{-G_B}{4(E_1K_1 + E_2K_2)},$$

$$\text{so that } M'_{B-} = -\frac{E_1K_1G_B}{E_1K_1 + E_2K_2}, \quad M'_{B+} = \frac{E_2K_2G_B}{E_1K_1 + E_2K_2}.$$

The new bending moments are

$$M_{B-} + M'_{B-} = -\frac{E_1K_1G_B}{E_1K_1 + E_2K_2} = \frac{E_2K_2M_{B-} + E_1K_1M_{B+}}{E_1K_1 + E_2K_2}$$

and

$$M_{B+} + M'_{B+} = M_{B+} + \frac{E_2K_2(M_{B-} - M_{B+})}{E_1K_1 + E_2K_2} = \frac{E_2K_2M_{B-} + E_1K_1M_{B+}}{E_1K_1 + E_2K_2},$$

and these are equal, so that the bending moment is continuous at B. No external moment is now required at B, but additional external moments are required at the other ends of the sections. The additional bending moment at  $A_+$  is  $M'_{A+} = -\frac{1}{2}M'_{B-}$ , while that at C is  $M'_{C-} = -\frac{1}{2}M'_{B+}$  and correspondingly increased external moments will have to be applied at A and C. These increases will be  $G'_A = -M'_{A+} = \frac{1}{2}M'_{B-}$  and

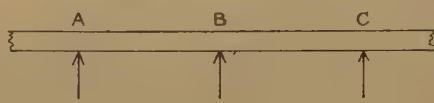
$$G'_C = +M'_{C-} = -\frac{1}{2}M'_{B+}.$$

The sum of these is  $G'_A + G'_C = \frac{1}{2}(M'_B - M'_{B+})$   
 $= -\frac{1}{2}G.$

Thus the original external moment has been removed and additional external moments totalling half this amount have been introduced. If this process is repeated at each junction in turn the sum of the external moments present will be progressively reduced until it becomes negligibly small. Thus the successive solutions converge towards the true solution for the original loading, the process being terminated when the additional moments have become negligible.

The factors  $\frac{E_1 K_1}{E_1 K_1 + E_2 K_2}$  and  $\frac{E_2 K_2}{E_1 K_1 + E_2 K_2}$   
 reduce to  $\frac{K_1}{K_1 + K_2}$  and  $\frac{K_2}{K_1 + K_2}$

Fig. 6.

						
DISTRIBUTION FACTORS	....	....	$f_1 = \frac{E_1 K_1}{E_1 K_1 + E_2 K_2}$	$f_2 = \frac{E_2 K_2}{E_1 K_1 + E_2 K_2}$	....	....
FIXED END MOMENTS	$M_{A-}$	$M_{A+}$	$M_{B-}$	$M_{B+}$	$M_{C-}$	$M_{C+}$
EXTERNAL MOMENTS			$M_{B-} - M_{B+}$ $= G_B$			
BALANCING MOMENTS			$-f_1 G_B$	$+f_2 G_B$		
TRANSFERRED MOMENTS			$\frac{1}{2}f_1 G_B$		$-\frac{1}{2}f_2 G_B$	
TOTAL MOMENTS						

TOTAL MOMENTS ARE OBTAINED BY SUMMING COLUMNS AND THE PROCESS TERMINATES WHEN THE EXTERNAL MOMENTS ARE ALL NEGLIGIBLE.

if  $E_1 = E_2$ , as is usual. These are known as the distribution factors for the junction B. The only restrictions on the choice of sections is that  $EI$  must be constant along the section so that the  $K$ 's are definite quantities for each section; and the end points of the sections must be rigid supports so that the external forces introduced are absorbed into the reactions of the supports.

The calculations for this method are set out in tabular form, a portion of such a table being shown in fig. 6. Of course, in a numerical



problem the actual figures would be inserted in the table. It is not the present purpose, however, to discuss this method in detail \*, but merely to show that the sign conventions and methods discussed in this paper are quite adequate to deal with this type of problem.

## APPENDIX I.

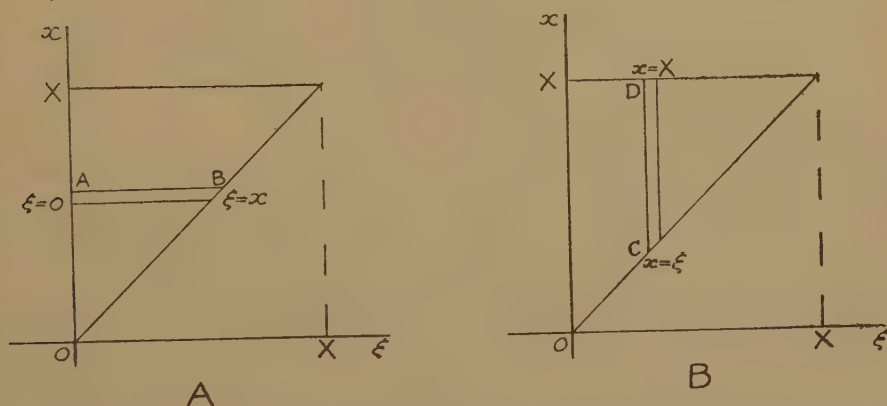
$$\int_0^X f(\xi) \frac{(x-\xi)^n}{n!} d\xi$$

is a definite integral with respect to  $\xi$ , which symbol will, therefore, disappear when the integral is evaluated and the limits inserted, but as  $x$  appears in the integrand as well as in the upper limit, the result of the integration, during which  $x$  is treated as a constant, will depend on the particular value given to  $x$ , *i. e.* it will be a function of  $x$ . Consider an integral of this type between limits 0 and  $X$ ,

$$\text{i. e.} \quad \int_0^X dx \int_0^x f(\xi) \frac{(x-\xi)^n}{n!} d\xi.$$

This may be evaluated directly by carrying out the two successive integrations, but it may be convenient to interchange the order of the two integrations. To effect this we regard  $(\xi, x)$  as co-ordinates of a point on a  $\xi x$  plane and convert the repeated integral into a double integral over the triangular region shown in fig. 7 A, the integration with respect

Fig. 7.



to  $\xi$  being along the strip AB from  $\xi=0$  to  $\xi=x$ , the contributions from these strips then being summed from  $x=0$  to  $x=X$  in the second integration. The double integral may now be re-arranged as a repeated integral, as shown in fig. 7 B, the summation being first taken along the strip CD from  $x=\xi$  to  $x=X$  with respect to  $x$  with  $\xi$  constant, the contributions of these strips being then summed from  $\xi=0$  to  $\xi=X$ .

\* Hardy Cross, Amer. I. C. E. (1930).

This gives

$$\begin{aligned}\int_0^X dx \int_0^x f(\xi) \frac{(x-\xi)^n}{n!} d\xi &= \iint f(\xi) \frac{(x-\xi)^n}{n!} dx d\xi \\ &= \int_0^X d\xi \int_\xi^X f(\xi) \frac{(x-\xi)^n}{n!} dx.\end{aligned}$$

Since  $f(\xi)$  is constant during the first integration, we obtain, on the right,

$$\begin{aligned}&\int_0^X d\xi f(\xi) \left[ \frac{(x-\xi)^{n+1}}{(n+1)!} \right]_\xi^X \\ \text{or } &\int_0^X d\xi f(\xi) \frac{(X-\xi)^{n+1}}{(n+1)!}.\end{aligned}$$

Thus, finally,

$$\int_0^X dx \int_0^x f(\xi) \frac{(x-\xi)^n}{n!} d\xi = \int_0^X f(\xi) \frac{(X-\xi)^{n+1}}{(n+1)!} d\xi,$$

in which the integration with respect to  $x$  is completed, but the result still left as an integral with respect to  $\xi$ .

## APPENDIX II.

The general formula for differentiating an integral

$$F(x) = \int_a^b f(x, \xi) d\xi$$

with respect to  $x$  when the limits  $a$  and  $b$  are variable is

$$F'(x) = \int_a^b \frac{\partial f(x, \xi)}{\partial x} d\xi + f(x, b) \frac{db}{dx} - f(x, a) \frac{da}{dx}.$$

In the case of the integral

$$F(x) = \int_0^x f(\xi) \frac{(x-\xi)^{n+1}}{(n+1)!} d\xi$$

we have  $a=0$ , so that  $\frac{da}{dx}=0$  and  $b=x$ , so that  $f(x, b)=0$  and only the first term on the right survives, giving

$$F'(x) = \int_0^x f(\xi) \frac{(x-\xi)^n}{n!} d\xi.$$

## LXXXVII. A Problem in Dynamic Balancing of Scanners for Radar.

By C. Fox, M.A., D.Sc.\*

[Received January 8, 1946.]

### § 1. Introduction.

In the first application of centimetric radar to the equipment of night fighters it was necessary to provide a scanner which projected the radio beam in a repeating spiral. The paraboloid diameter for the reflector was 28 in. in diameter, the rotation of the axis of the paraboloid about

\* Communicated by the Author.

an axis fixed in the aircraft was 1000 r.p.m. and the repeat period of scan was 1.2 seconds. The maximum angle of tilt between the paraboloid and the axis of rotation was  $30^\circ$ .

Figs. 1 and 2 show the complete scanner assembly with the various parts numbered as follows:—On a base plate 1 is mounted a cast hub 2 containing bearings within which rotate two concentric hollow shafts which are driven by chains and sprockets through a gear box 3 by a hydraulic motor 4. The speed of the motor is controlled by a centrifugal governor operating a hydraulic valve 5.

The outer hollow shaft carries a cast yoke 6, with bearings 7 perpendicular to the axis of rotation, and in these bearings is supported a casting forming part of the paraboloid assembly 8. The inner hollow shaft carries a roller 9 disposed so that it is eccentric to and its axis inclined to the axis of the hollow shafts. The axes of rotation of the hollow shafts, of the bearings supporting the paraboloid assembly on the yoke, and of the eccentric roller all intersect at a point (O in fig. 3). The roller operates, with flat bearing plates 10, on the paraboloid assembly which define its tilt relative to the yoke. The system is dynamically balanced by two weights one of which is indicated by W.

The outer hollow shaft is driven at 1000 r.p.m. and the inner hollow shaft at 938 r.p.m., thus causing the paraboloid assembly to rotate and tilt relative to its axis of rotation at the required speeds.

To avoid excessive bearing and structural loads and vibration, it was necessary to secure dynamic balance at all inclinations of the paraboloid assembly, and to enable this to be done the conditions of balancing such a dynamic system were investigated and a general method of balancing determined. This article describes the mathematical reasoning and the form in which it was used for design purposes.

## § 2. *Scanner Description.*

Fig. 3 illustrates, diagrammatically, the mounting of the reflector as seen from behind. The two concentric circles represent the reflector, ABCD represents the yoke which rotates with constant angular velocity  $\omega$  about the axis EF, and AB is the oscillation axis for the reflector and those components of the mechanism which are rigidly attached to it (*e.g.* the two weights W and W').

The whole of the mechanism which oscillates about AB will for the rest of the article be referred to as the reflector assembly.

Fig. 4 is a cross-section of the apparatus perpendicular to AB. The parabolic arc represents the reflector, O is the intersection of EF and AB (which is perpendicular to the plane of the diagram), and OH is any line fixed in the reflector assembly and perpendicular to AB. If the axis of the reflector passes through O and is at right angles to AB it may be conveniently taken for the line OH. W and W' represent two balance weights.

The oscillation of the reflector assembly about AB can evidently be measured in terms of the angle  $\theta$  between EF and OH, and the motion is

Fig. 1.

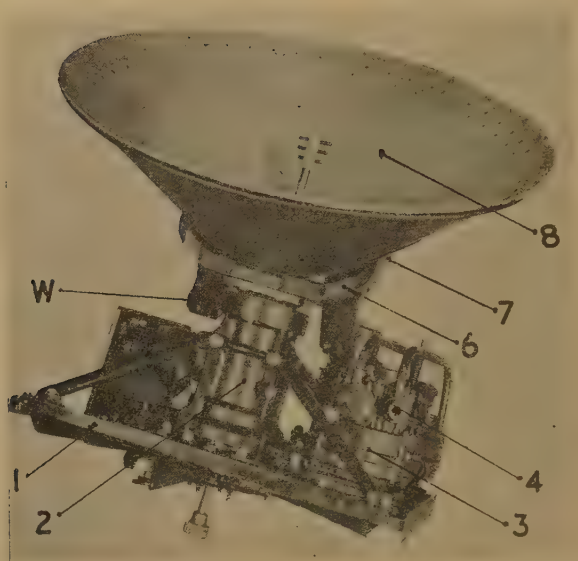
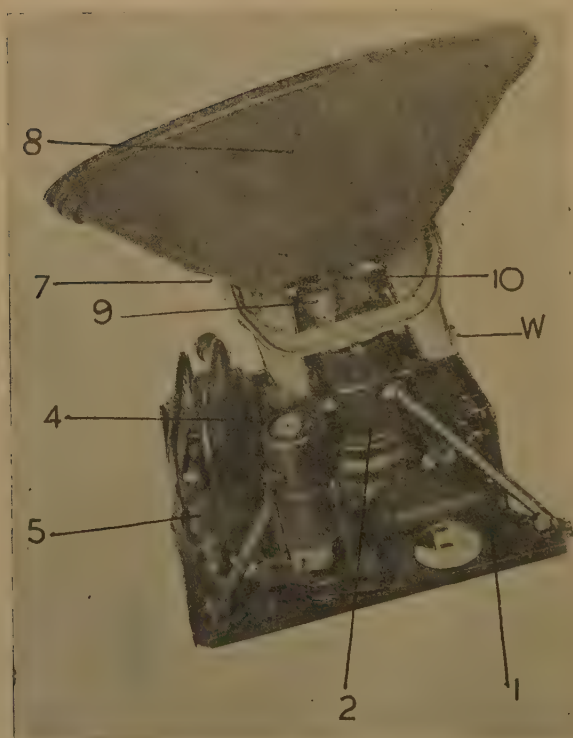


Fig. 2.





analogous to that of a spinning top with  $\dot{\theta}$  as nutation. The spin  $\omega$  about EF and the nutation  $\dot{\theta}$  about AB are not independent of each other; in most cases they are generated by the same motor and the mechanism of the scanner can be designed so that the relationship between them satisfies any prescribed condition.

Fig. 3.

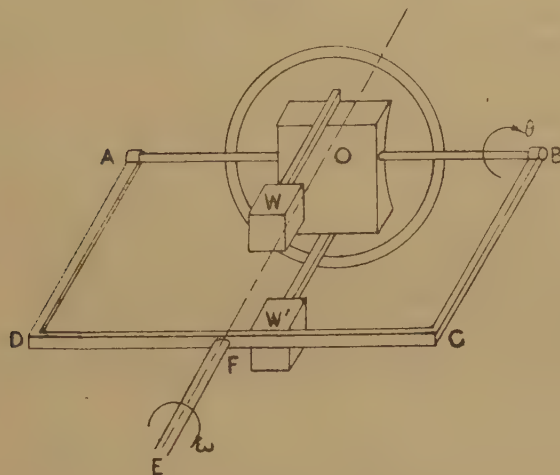
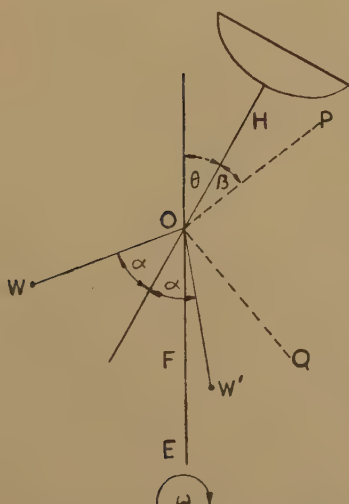


Fig. 4.



For example, in the spiral scanner under consideration, the relation is

$$\dot{\theta} = c\omega \cos^2 \theta \sqrt{(\tan^2 \lambda - \tan^2 \theta)}, \quad \dots \dots \dots (1)$$

where  $c$  and  $\lambda$  are constants. The constant  $c$  is so small in practice that the magnitude of  $\dot{\theta}$  does not exceed about 3 per cent. that of  $\omega$ , and the constant  $\lambda$ , which evidently measures the extreme angle of oscillation, can be in some cases as much as  $30^\circ$  or so.

§ 3. *Conditions for Dynamic Balance.*

By means of Euler's equations \* it will now be shown that the necessary and sufficient conditions for dynamic balance of the reflector assembly are as follow :—

- (i) The centre of gravity of the reflector assembly must be at O.
- (ii) AB must be a principal axis of inertia at O for the reflector assembly.
- (iii) The moments of inertia of the reflector assembly about the two principal axes of inertia through O perpendicular to AB must be equal to each other.

The proof is in two parts : the first part, showing that the conditions are sufficient, is comparatively simple, since AB can then be assumed to be a principal axis. The second part, showing that the conditions are necessary, is more involved and will be relegated to an appendix at the end of the article. The practical application of the theory to balance problems will follow immediately after the first part.

§ 4. *Sufficiency of Balance Conditions.*

Condition (i), § 3, is almost self-evident. If the centre of gravity G is not at O, let its distance from EF at any time  $t$  be  $z$  and let  $M$  be the mass of the reflector assembly. Then G has an inward centrifugal acceleration of magnitude  $z\omega^2$  and the forces parallel to this direction acting on the assembly must have a resultant of magnitude  $Mz\omega^2$  acting at G. At 1000 r.p.m.  $\omega=100\pi/3$ , and so this force will be extremely large. Therefore, for good dynamic balance it must be eliminated and the elimination can be effected for all values of  $t$ , only if  $z=0$ , i. e. if G is at O.

Assuming the truth of condition (ii) in § 3 that AB is a principal axis of inertia, let OP and OQ be the other two principal axes at O (they are shown as discontinuous lines in fig. 4). Since OP and OQ are perpendicular to AB, they will have angular velocity  $\dot{\theta}$ , and so at time  $t$  the angle between EF and OP will be  $\theta+\beta$ , where  $\beta$  denotes the constant angle HOP.

Taking the positive direction of OA as the one for which a clockwise rotation through  $90^\circ$  turns OP to OQ, and the positive directions of OP and OQ as shown in fig. 4, then OA, OP, OQ, in this order, form a right-handed system of axes. The components of the two angular velocities  $\dot{\theta}$  about AB and  $\omega$  about EF along these axes are evidently

$$\{\dot{\theta}, \omega \cos (\theta+\beta), -\omega \sin (\theta+\beta)\}, \quad . \quad . \quad . \quad . \quad (2)$$

respectively. If  $I_a, I_p, I_q$  are the moments of inertia of the reflector assembly about OA, OP and OQ respectively, then, since they are principal axes, the components of angular momentum about these axes are

$$\{I_a \dot{\theta}, I_p \omega \cos (\theta+\beta), -I_q \omega \sin (\theta+\beta)\}. \quad . \quad . \quad . \quad . \quad (3)$$

---

\* Routh "Rigid Dynamics," Whitaker "Analytical Dynamics," and similar works on dynamics.

Finally, if  $(\mu_a, \mu_p, \mu_g)$  denote the moments about these axes of the couples required to generate the motion of the assembly, Euler's equations of motion state that

$$I_a \ddot{\theta} - I_g \omega^2 \sin(\theta + \beta) \cos(\theta + \beta) + I_p \omega^2 \sin(\theta + \beta) \cos(\theta + \beta) = \mu_a, \quad (4)$$

$$-I_p \omega \dot{\theta} \sin(\theta + \beta) - I_a \dot{\theta} \omega \sin(\theta + \beta) + I_g \dot{\theta} \omega \sin(\theta + \beta) = \mu_p, \quad (5)$$

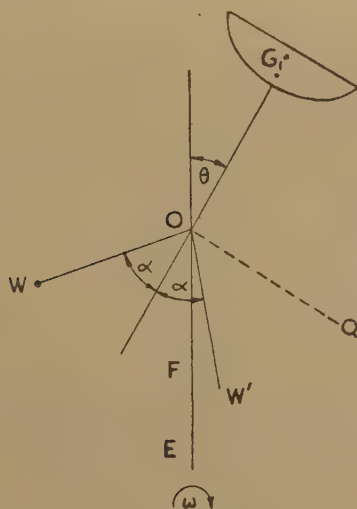
$$-I_g \omega \dot{\theta} \cos(\theta + \beta) + I_p \dot{\theta} \omega \cos(\theta + \beta) - I_a \dot{\theta} \omega \cos(\theta + \beta) = \mu_g. \quad (6)$$

Evidently the best running conditions occur when these couples are as small as possible. Since  $\dot{\theta}$  and  $\ddot{\theta}$  are small compared to  $\omega$  (generally less than 3 per cent. of it), the largest terms of the equations will be those involving  $\omega^2$ . Such terms occur in equation (4) and will only cancel if  $I_g = I_p$ , thus proving the sufficiency of conditions (ii) and (iii), § 3, for dynamical balance.

### § 5. Balancing by Means of Two Weights.

In practice, reflector assemblies do not usually satisfy conditions (i) and (iii), § 3. But in all cases AB and  $OG_1$ , where  $G_1$  denotes the centre of

Fig. 5.



gravity of the assembly, are perpendicular to each other and are principal axes of inertia (usually because  $ABOG_1$  is a plane of symmetry). Calling such assemblies dynamically unbalanced, it can be shown from the preceding theory that they can be dynamically balanced by the addition of two weights. The two weights can be merged into one if convenient.

Fig. 5 shows a cross-section of an unbalanced assembly perpendicular to AB together with two balance weights. The reflector is represented by a parabolic arc, as in fig. 4;  $G_1$  is the centre of gravity of the assembly without the balance weights, and  $W$  and  $W'$  are the centres of gravity of

the two balance weights. The principal axes of inertia at O are OA (perpendicular to the plane of the diagram),  $OG_1$  and OQ (shown as a discontinuous line). The angle  $WOW'$ , which can be chosen at will, is equal to  $2\alpha$  and is bisected by  $G_1O$  produced. The difference between figs. 4 and 5 is that the lines OP and OH of fig. 2 both coincide with  $OG_1$  in fig. 5.

Let  $OG_1=l$ , where  $l$  is positive if  $G_1$  is on the side of O marked in the diagram and negative on the other side. Let  $OW=OW'=d$  and let  $M$  denote the mass of the unbalanced assembly and  $m$  that of each balance weight. Then, if the addition of these weights results in a balanced assembly, condition (i), § 3, states that the centre of gravity of the complete system must be at O. Taking moments about O, it follows that

$$2md \cos \alpha = ML \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Let  $I_g$  and  $I_t$  denote the moments of inertia of the unbalanced assembly about  $OG_1$  and the transverse axis OQ respectively. Let  $i_g$  and  $i_t$  denote the moments of inertia of each balance weight about axes through its centre of gravity parallel to  $OG_1$  and OQ respectively, then about  $OG_1$  and OQ the moments of inertia will be  $i_g+md^2 \sin^2 \alpha$  and  $i_t+md^2 \cos^2 \alpha$  respectively. Hence, from condition (iii), § 3, it follows that dynamical balance can be attained if

$$2(i_g+md^2 \sin^2 \alpha)+I_g=2(i_t+md^2 \cos^2 \alpha)+I_t \quad . \quad . \quad . \quad (8)$$

Now it is possible for the weights to have such a shape that

$$i_g=i_t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

in which case, equation (8) simplifies. For example, if the weights are cubes with one of their faces perpendicular to AB, then  $i_g=i_t$ . Since weights used in practice always satisfy this condition (they were generally cubes), equation (8) then simplifies to

$$2md^2 \sin^2 \alpha + I_g = 2md^2 \cos^2 \alpha + I_t \quad . \quad . \quad . \quad . \quad (10)$$

In other words, for such shapes the weights may be treated dynamically as particles having zero moment of inertia.

On solving equations (7) and (10) for  $d$  and  $m$ , we have

$$d = \frac{(I_g - I_t) \cos \alpha}{Ml \cos 2\alpha}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$m = \frac{M^2 l^2 \cos 2\alpha}{2(I_g - I_t) \cos^2 \alpha} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Hence, on measuring  $M$ ,  $l$ ,  $I_g$  and  $I_t$  (the last two quantities by bifilar suspension or any similar method), the position and mass of the balance weights can be found from equations (11) and (12) and the balance problem solved.





The arguments just used prove the following result :—Denote by  $J$  a small part of the assembly in the immediate neighbourhood of a point  $X$ , which is situated in the plane  $G_1OQ$ . Then, if the matter of  $J$  is symmetrically distributed about  $X$  along a line parallel to  $AB$ , the value of  $I_g - I_t$  and the position of the centre of gravity of the assembly remain unaltered. The dynamic balance of the system is therefore not affected by such redistribution.

Equations (11) and (12) fail to give solutions in the three following cases :—

- (i)  $I_g = I_t$  and  $l = 0$ .
- (ii)  $I_g = I_t$  and  $l \neq 0$ .
- (iii)  $I_g \neq I_t$  and  $l = 0$  (*i. e.*,  $G_1$  is on  $AB$ ).

In case (i) the assembly is already balanced and needs no further consideration.

In case (ii) weights can be attached to the unbalanced assembly so as to break down the relation  $I_g = I_t$  and the system can then be balanced by the previous theory. A better method is as follows :—Reverting to fig. 5, consider lines  $OW$  and  $OW'$  for which  $\alpha = 45^\circ$ . Since weights placed on such lines are equidistant from  $OG_1$  and  $OQ$ , they will have equal moments about these axes and so cannot affect the difference  $I_g - I_t$ . If there are two such weights, one on  $OW$  and the other on  $OW'$ , each of mass  $m_2$  and distant  $d_2$  from  $O$ , then the centre of gravity of the assembly together with the weights will be at  $O$  if

$$2m_2d_2 \cos 45^\circ = Ml, \quad \dots \dots \dots (15)$$

where, as before,  $M$  is the mass of the unbalanced assembly and  $OG = l$ .

Equation (15) ensures the satisfaction of condition (i), § 3, and the relation  $I_g = I_t$ , which still holds, ensures the satisfaction of condition (iii), § 3. Evidently  $OA$ ,  $OG_1$  and  $OQ$  still remain principal axes, so ensuring the satisfaction of (ii), § 3. The system is therefore dynamically balanced.

This case has one advantage since the value of  $d_2$  can be chosen at will.

Case (iii) can also be dealt with by attaching auxiliary weights to the unbalanced assembly so as to break down the relation  $l = 0$  and then balancing according to the previous theory. Another method is to place weights along the axis  $OQ$  symmetrically about  $O$ . Evidently  $l = 0$  still holds and so (i), § 3, is satisfied. If there are two weights each of mass  $m_3$  and distant  $h_3$  from  $O$ , then condition (iii), § 3, will be satisfied if

$$2m_3h_3^2 + I_g = I_t. \quad \dots \dots \dots (16)$$

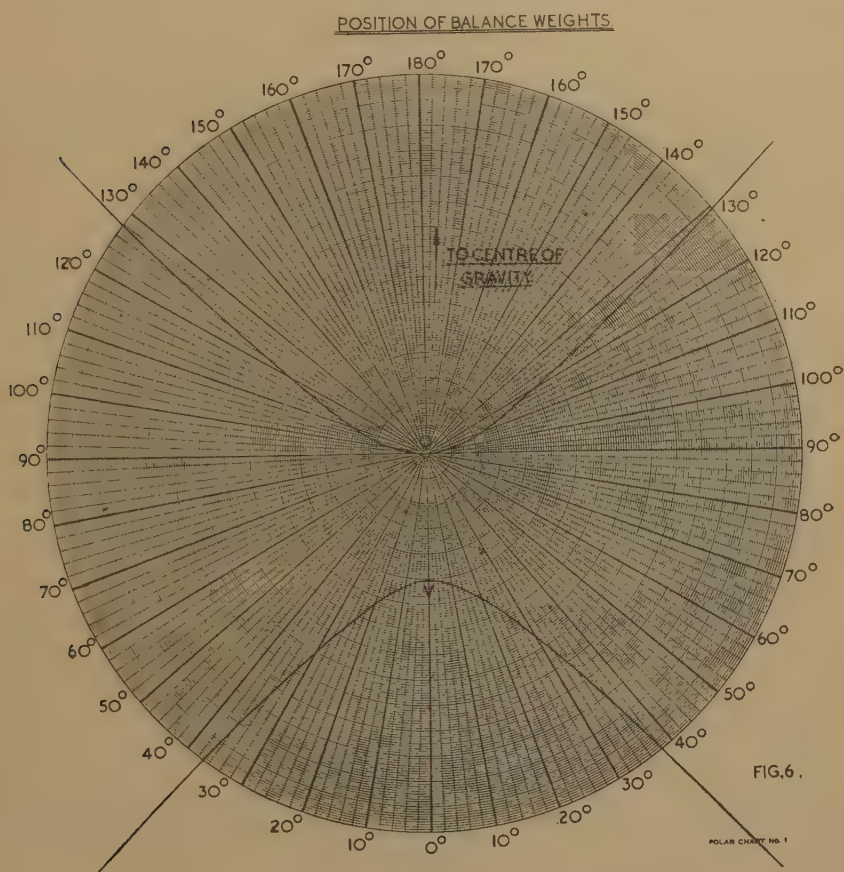
As it is evident that  $OA$ ,  $OG_1$  and  $OQ$  still remain principal axes, it follows that the system will be dynamically balanced. If  $I_g < I_t$ , then  $m_3$  is positive, and if  $I_g > I_t$ , then  $m_3$  is negative. A negative value of  $m$  means that metal must be removed from the axis  $OQ$ , a course of action which is usually impracticable.

If weights can be placed on OQ, then case (iii) offers a convenient method of balancing in practice. The assembly is first balanced statically and then balanced dynamically by means of equation (16). Its advantages are that statical balance is comparatively easy to obtain and that the value of  $h_3$  in equation (16) can be chosen at will.

### § 8. Graphical Representations for Equations (11) and (12).

Taking  $d$  and  $\alpha$  as polar co-ordinates, with  $d$  as the radius vector and  $\alpha$  as the vectorial angle, fig. 6 represents equation (11) graphically. On

Fig. 6.



transforming the equation into cartesian co-ordinates by using the substitutions  $x=d \cos \alpha$  and  $y=d \sin \alpha$ , equation (11) becomes

$$x^2 - y^2 = \frac{I_g - I_t}{Ml} x, \quad \dots \dots \dots (17)$$

which is a rectangular hyperbola. The graph is used as follows:— Determine the value of  $(I_g - I_t)/Ml$  by experiment, choose the value of  $\alpha$



and then draw a line through O making an angle  $\alpha$  with the line OV on the graph. If this line cuts the hyperbola at the point Z, then

$$d = \frac{OZ}{OV} \frac{(I_g - I_t)}{Ml} \dots \dots \dots (18)$$

Taking  $m$  and  $\alpha$  as polar co-ordinates,  $m$  being the radius vector and  $\alpha$  the vectorial angle, fig. 7 represents equation (12) graphically. To find  $m$

Fig. 7.

MASS OF BALANCE WEIGHTS

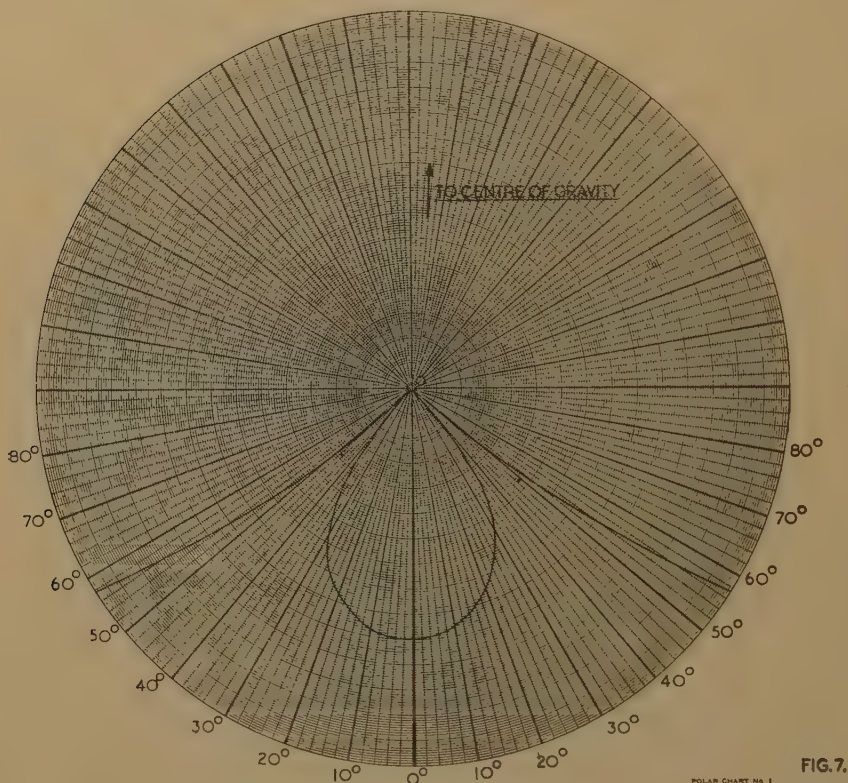


FIG. 7.

from the graph, through O draw a line making an angle  $\alpha$  with  $00^\circ$  and let this line meet the curve at Y. Then

$$m = \frac{OY \cdot M^2 l^2}{OU \cdot 2(I_g - I_t)} \dots \dots \dots (19)$$

### § 9. Tests for Balance.

The best test of dynamic balance is that the assembly should operate at a high speed of rotation with negligible vibration. There is, however, another test which can be sometimes used and which is cognate to the



subject. If the assembly is statically balanced and then mounted so that it can swivel freely about the axis AB without any couple being exerted on it, equation (4) becomes

$$I_a \ddot{\theta} - I_t \omega^2 \sin \theta \cos \theta + I_g \omega^2 \sin \theta \cos \theta = 0, \quad \dots \quad (20)$$

where  $I_p$  has been replaced by  $I_g$ ,  $I_g$  by  $I_t$ , and  $\beta$  has been taken as zero in order to conform with the notation of § 5 and the lettering of fig. 5.

If the system is balanced, *i. e.*,  $I_g = I_t$ , then  $I_a \ddot{\theta} = 0$ , and so  $\dot{\theta}$  is constant. Hence if initially  $\dot{\theta} = 0$ , then the axis of the reflector will make a constant angle with the shaft EF throughout the motion.

If  $I_g < I_t$ , then  $\ddot{\theta}$  is positive and  $\dot{\theta}$  will steadily increase. Hence the axis of the reflector will swing out until ultimately it is at right-angles to the shaft EF.

If  $I_g > I_t$ , then  $\ddot{\theta}$  is negative and  $\dot{\theta}$  will steadily decrease and the axis of the reflector will swing inwards until it is in line with the shaft EF.

Tests of this kind made on models confirm the theory completely.

### § 10. Appendix.

The theory will now be completed by showing that conditions (ii) and (iii), § 3, are necessary as well as sufficient for dynamic balance. The necessity as well as the sufficiency of condition (i) have been shown in § 4.

The assumption that OA, OP and OQ (fig. 4) are principal axes is now no longer admissible, and so the products of inertia must be taken into account.

Let these products be denoted by  $P_a$ ,  $P_p$  and  $P_g$ . As in § 4, the moments of inertia will be denoted by  $I_a$ ,  $I_p$  and  $I_g$ .

The components of  $\dot{\theta}$  about AB and of  $\omega$  about EF along the three axes OA, OP and OQ (fig. 4) are as before :

$$\{\dot{\theta}, \omega \cos (\theta + \beta), -\omega \sin (\theta + \beta)\}, \quad \dots \quad (20)$$

and for brevity these will be denoted by  $(\omega_1, \omega_2, \omega_3)$ . The corresponding components of angular momentum about these axes are now

$$\{(I_a \omega_1 - P_g \omega_2 - P_p \omega_3), (I_p \omega_2 - P_a \omega_3 - P_g \omega_1), (I_g \omega_3 - P_g \omega_1 - P_a \omega_2)\}, \quad (21)$$

which for brevity will be denoted by  $(k_a, k_p, k_g)$ . If, as in § 4,  $(\mu_a, \mu_p, \mu_g)$  denote the moments of the couples about OA, OP, OQ respectively required to generate the motion, then Euler's equations state that

$$\left. \begin{aligned} \dot{k}_a + \omega_2 k_g - \omega_3 k_p &= \mu_a, \\ \dot{k}_p + \omega_3 k_a - \omega_1 k_g &= \mu_p, \\ \dot{k}_g + \omega_1 k_p - \omega_2 k_a &= \mu_g \end{aligned} \right\} \dots \quad (22)$$

Now, the best running conditions occur when the magnitudes of these couples are as small as possible. But the magnitudes of  $\dot{\theta}$  and  $\ddot{\theta}$  are, in practice, less than 3 per cent. that of  $\omega$ , and so the largest terms in

equations (22) will be those containing  $\omega^2$ . Hence the best conditions will be obtained if the coefficients of the terms in  $\omega^2$  can be made equal to zero. From equations (20) and (22) it can be seen that the terms involving  $\omega^2$  are as follows :—

$$\text{in } \mu_a \text{ they are } I_g \omega_2 \omega_3 - P_a \omega_2^2 - I_p \omega_2 \omega_3 + P_a \omega_3^2,$$

$$\text{in } \mu_p \text{ they are } -P_g \omega_2 \omega_3 - P_p \omega_3^2,$$

$$\text{in } \mu_g \text{ they are } P_g \omega_2^2 + P_p \omega_2 \omega_3,$$

where  $\omega_2 = -\omega \sin(\theta + \beta)$  and  $\omega_3 = \omega \cos(\theta + \beta)$ . Evidently these terms can vanish for all values of  $\theta$  only if  $P_a = 0$ ,  $P_p = 0$ ,  $P_g = 0$  and  $I_p = I_g$ , thus proving that conditions (ii) and (iii), § 3, are necessary as well as sufficient for dynamic balance.

### § 11. *Summary and Acknowledgments.*

This article deals with the balancing of a scanner having a spiral scan which was developed and designed by Messrs. Nash & Thompson, Ltd., under the direction of Mr. A. Whitaker, O.B.E., M.A., F.Inst.P., M.I.E.E., during the latter part of 1940.

## LXXXVIII. *The Second Law of Thermodynamics, with Particular Reference to Carnot Theorems Relating to Regenerative, Refrigeration and Warming Cycles.*

By Professor W. J. WALKER, Department of Mechanical Engineering, University of the Witwatersrand, Johannesburg, South Africa\*.

[Received June 18, 1946.]

### PRÉCIS.

The following communication sets out, at first, to demonstrate that the so-called self-evident proposition adopted by Ostwald as his statement of the Second Law of Thermodynamics, and by Planck in his derivation of that law in terms of entropy, need not be assumed as self-evident, since it can be proved on lines following those of Carnot's proof of his theorem. There would appear to be no more justification for assuming the Ostwald-Planck proposition as self-evident than there would be in accepting the Carnot Heat Engine Theorem without proof.

The remainder of the paper is devoted to an elucidation of the problem of determining the comparable Carnot temperature limits to be selected for any irreversible engine or refrigerating cycle, and, after developing the expressions for efficiency and coefficient of performance for engine and refrigerator regenerative cycles respectively, the warming efficiency of combinations of regenerative engine and refrigeration cycles is finally examined.

\* Communicated by the Author.

It is shown that the determination of the comparable Carnot temperature limits is always unique for either engine or refrigeration cycles, although the method of determination is different for the two types. An expression for warming efficiency is obtained in terms of the thermal efficiency of the engine cycle and the coefficient of performance of the refrigerator cycle.

The paper closes with a reference to the possibilities of the regenerative cycle in relation to the general problem of the cooling of hot mine air.

---

### 1. Introduction.

THE main purpose of the following contribution is to make clear the implications of the Carnot principle to regenerative, refrigeration and warming cycles, since it does not appear that the definition of temperature limits involved in the principle has been made sufficiently explicit to remove the danger of deducing false conclusions by the adoption of wrong comparable Carnot temperature limits and, more important still, of stultifying the evolution of more efficient refrigeration and warming cycles. It will appear hereafter that, fortunately enough, for heat engine cycles the choice of such temperature limits has been correctly made as the maximum and minimum temperatures of the cycle under examination, although no rigid proof for the necessity of such a choice has apparently ever been given. The result, however, has been a much greater development in variety of heat engine cycles, both in theory and practice, than is the case for either refrigerator or warming cycles.

At the same time, an attempt has here been made, at the outset of the analysis, to develop a more logical rendering of the interrelations between the Second Law of Thermodynamics and the Carnot principle. It appears that this can be done by basing all deductions on Clausius' definition of the Second Law, namely, that "Heat cannot, of itself, without other changes in its environment, flow from a container of lower temperature to one of higher temperature." The term "flow," as here used, includes the conceptions of transfer, not only by conduction but any other means. It would appear that the more usual approach to this law, as outlined in the methods of Lord Kelvin, Clerk Maxwell, Ostwald and Planck, does not provide that single line of logical development so necessary to obviate either a retracing of steps in the argument, or, what is even less desirable, a recourse to self-evidency or *à priori* characteristics in some given statement. If one is satisfied with Clausius' statement of the Second Law, and, from the macroscopic point of view, it is difficult to see any fundamental objection to it, it is here shown that it includes all the other correlated statements which appear to have been made, whether derived or based on so-called self-evident propositions.

Clerk Maxwell, following Kelvin's demonstration, employs, in his proof of the Carnot principle, the axiom that it is impossible to convert the heat of the colder of two reservoirs in a system into mechanical work. Strangely enough, neither Kelvin nor Maxwell appear to have recognized



that, following the same line of reasoning which they employed, Carnot's principle can also be proved by laying down the axiom that it is impossible, without contemporaneous alteration in the heat content of the colder of two reservoirs in a system, to convert the heat of the warmer reservoir into mechanical work. In these axioms, of course, it is assumed that, in the first, the heat of the warmer reservoir, and, in the second, the heat of the colder reservoir, is maintained constant throughout.

It would appear that it is this omission which, unobserved by Planck, has led the latter, in his *Treatise on Thermodynamics*, to develop the theory of entropy, on which he finally bases his statement of the Second Law, by utilizing at the outset what he terms a "self-evident proposition," which he states as follows:—"It is impossible to construct an engine which will work in a complete cycle and produce no effect except the raising of a weight and the cooling of a heat reservoir." It will be observed that Planck here makes no reference to the relative temperature level of the heat reservoir, so that he is assuming, as an axiom, a wider generalization than that of Kelvin. It is to be noted also, however, that Kelvin's statement is applied to a proof of Carnot's theorem, whereas that of Planck is used to develop the entropy equivalent of what he considers to be the most general statement of the Second Law. Planck's self-evident proposition is more or less a paraphrase of Ostwald's statement of the Second Law, namely, "That perpetual motion of the second type is impossible, even with ideal machines, *i. e.*, even with frictionless, heat-insulated engines," perpetual motion of the second type being defined as "the transformation of heat into work, without changing anything else in the environment." Having accepted this so-called self-evident proposition, Planck then arrives at what he considers to be the most general statement possible of the Second Law, namely, "Every physical or chemical process in nature takes place in such a way as to increase the sum of the entropies of all the bodies taking any part in the process. In the limit, *i. e.* for reversible processes, the sum of the entropies remains unchanged." This entropy statement, while substantially similar to that of Clausius bearing on the interpretation of the entropy function in physical processes, is capable of wider application to chemical as well as physical problems in general, and, in the hands of Planck, its universality is made clear. To those who appreciate its value, therefore, it appears a weakness that the whole of this wonderful edifice should be made to rest on a so-called self-evident proposition, the substantiation of which is not at all convincingly demonstrated by Planck, nor, so far as can be ascertained, by anyone else. While the conception of entropy, as a fundamental quantity in thermodynamic analysis, has been of inestimable value in such analysis, it is difficult to agree that its character, as *indirectly* displayed in all natural phenomena, should be deemed to emphasize the universal tendency to macroscopic irreversibility of heat energy phenomena, more than the *direct* evidential character of heat flow itself, as set forth in the statement of the Second Law given by Clausius. As is well known, the proof of Carnot's heat engine theorem can be made to rest on Clausius' statement. Planck does not refer to this theorem of



Carnot's until he proves it from his own statement of the Second Law as based on the entropy conception, which, it should again be noted, depends on the formulation of an axiom, the self-evidency of which may, at least, be doubted. In the following analysis in the next section, using Clausius' statement of the Second Law, it will be shown that this self-evident proposition can be proved on lines following those adopted by Carnot, in proving his famous theorem.

## 2. Proof of the Ostwald-Planck Proposition.

Since the proof of this proposition, as given here, depends on the Carnot theorem, this is first given as follows :—

(a) *Proof of Carnot's Theorem.* Let there be two engines  $E_r$  (reversible) and  $E_i$  (irreversible) working between two temperature limits  $T_1$  and  $T_2$ . Let it be assumed that  $E_i$  has a higher efficiency than  $E_r$ , and that both engines do work at the same rate. Since, by hypothesis,  $E_i$  has a higher efficiency than  $E_r$ ,  $E_i$  will take less heat from the warmer reservoir. Let  $E_i$  drive  $E_r$  reversed. The result will be that the warmer reservoir will gain heat and no net mechanical work will be done. Since, by the First Law, the heat gained by the warmer reservoir can come only from the colder one, it follows that this result is a violation of the Second Law, as stated by Clausius. Hence an irreversible engine cannot be more efficient than a reversible one.

The proof of what is here called the Ostwald-Planck proposition will now be demonstrated by the enunciation and proof of two theorems, A and B.

(b) *Theorem A.* It is impossible, without contemporaneous alteration in the heat content of the colder of two reservoirs in a system, to convert the heat of the warmer reservoir into mechanical work.

Let there be two engines,  $E_r$  (reversible) and  $E_i$  (irreversible) working between temperatures  $T_1$  and  $T_2$ . Let it be assumed that the warmer reservoir can lose heat while net mechanical work is being done and that both engines discharge heat to the colder reservoir, each at the same rate,  $Q_2$ . Let  $E_i$  drive  $E_r$  reversed. If the warmer reservoir is to lose heat, the heat  $Q_1'$  withdrawn therefrom by  $E_i$  must be greater than  $Q_1$ , the heat discharged thereto by  $E_r$ . Also, since  $Q_1' - Q_2$  must be greater than  $Q_1 - Q_2$  under such circumstances, net mechanical work must be done at the same time. This can take place only if the efficiency of  $E_i$  is greater than that of  $E_r$ . But this, by Carnot's theorem, is impossible. Hence, it is impossible, without contemporaneous alteration in the heat content of the colder of two reservoirs in a system, to convert the heat of the warmer reservoir into work.

(c) *Theorem B.* It is impossible, with or without contemporaneous alteration in the heat content of the warmer of two reservoirs in a system, to convert the heat of the colder reservoir into mechanical work.

The slight addition to the form of this theorem, as compared with that of Theorem A, should be noted, since the proof of the first alternative of

Theorem B follows naturally from Clausius' definition of the Second Law. No such alternative applies in the case of Theorem A. The second alternative of Theorem B is proved as follows, on similar lines to the above.

Let there be two engines,  $E_r$  (reversible) and  $E_i$  (irreversible) working between temperatures  $T_1$  and  $T_2$ . Let it be assumed that the colder reservoir can lose heat while net mechanical work is being done, and that both engines discharge heat to the warmer reservoir, each at the same rate,  $Q_1$ . Let  $E_i$  drive  $E_r$  reversed. If the colder reservoir is to lose heat, the heat  $Q_2'$  discharged thereto by  $E_i$  must be less than that, say  $Q_2$ , withdrawn therefrom by  $E_r$ . Also, since  $Q_1 - Q_2'$  must be greater than  $Q_1 - Q_2$  under such circumstances, net mechanical work must be done at the same time. This can take place only if the efficiency of  $E_i$  is greater than that of  $E_r$ . But this, by Carnot's theorem, is impossible, hence, it is impossible, with or without contemporaneous alteration in the heat content of the warmer of two reservoirs in a system, to convert the heat of the colder reservoir into mechanical work.

From the foregoing proofs of theorems A and B, the Ostwald-Planck proposition naturally follows.

### 3. *General Considerations Relating to Comparable Carnot Temperature Limits.*

The method of proof used by Carnot can readily be applied to prove the statement that "No refrigerator operating between definite temperature limits can be more efficient than a reversible one operating between the same temperature limits." Ewing appears to have been the first to show this. It has not, apparently, however, been demonstrated what are the definite temperature limits in any comparison between an irreversible cycle and the Carnot cycle. True, as already stated, it is customary to take the maximum and minimum temperatures of the irreversible engine cycle under review as the comparable limits, but scientific justification for any such choice cannot be based on the principles of "right of way." In the case of the comparable limits for irreversible refrigeration cycles, no such definite practice appears to exist. To illustrate the point which is under consideration, figs. 1 and 2 have been drawn.

These two diagrams illustrate the Carnot cycle as an engine and a refrigerator cycle, respectively, and are drawn for the purpose of emphasizing that, in the limit, the infinite sources and sinks required for the ideal operation of the Carnot cycle require to have temperatures which are, differentially speaking, oppositely disposed for the two cycles in relation to the maximum and minimum temperatures of the Carnot cycle in each case. The direction of heat-flow from the source to the working fluid, and from the working fluid to the sink, in each case must clearly be the same, as indicated by the arrow-heads. This distinction, with reference to the comparable temperature limits, while implicit in the theory, does not appear to have been hitherto emphasized.

Referring now to fig. 3, this represents any irreversible engine cycle, the performance of which it is desired to compare with a Carnot cycle,

operating between the same temperature limits of source and sink. There is no *a priori* reason why one should not draw any two horizontal lines as AB and CD to represent the temperature levels of source and sink respectively. If, however, the conditions just after point A be examined, it will be observed that the heat content of the working medium has increased, whereas, as indicated by the arrow-heads, it is discharging heat to the source. Since the source is the only means of heat supply, such a

Fig. 1.

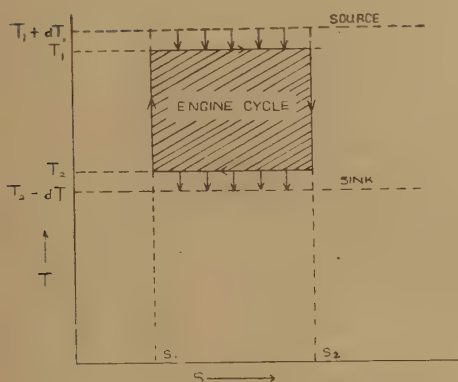


Fig. 2.

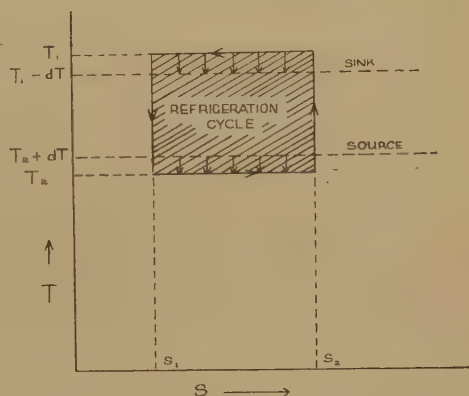
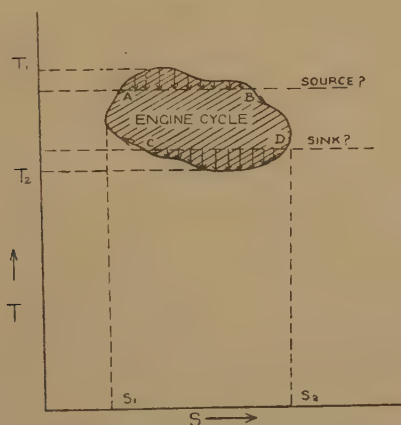


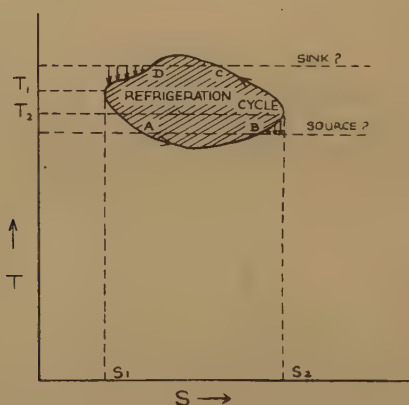
Fig. 3.



state of affairs is impossible. It follows, therefore, that AB must pass through  $T_1$ , the maximum temperature of the irreversible cycle. Similarly, considering the conditions just after point D, the heat content of the working medium is diminishing while heat is, as indicated by the arrow-heads, being received from the sink. Since the sink supplies the only means of heat withdrawal, such a state of affairs is again impossible. Hence, CD must pass through  $T_2$ , the minimum temperature of the irreversible cycle.

By the same method of reasoning, referring to the refrigerator cycle of fig. 4, the following result is deduced. Again, any two lines, AB and CD, are drawn to represent source and sink respectively. Considering the conditions after point B, the heat content of the working medium is increasing, while heat, as indicated by the arrow-heads, is being discharged to the source. Since the source provides the only means of supply of heat, this state of affairs is impossible; hence point B must be represented by the extreme right-hand point of the diagram, *i. e.* the line AB must pass through  $T_2$ . In the same way, considering conditions after point D, the heat content of the working medium is diminishing, while at the same time it is, as shown by the arrow-heads, drawing heat from the sink. Since the sink supplies the only means of heat withdrawal this state of affairs is impossible. Hence, point D must be represented by the extreme left-hand point of the diagram, *i. e.* the line DC must pass through  $T_1$ .

Fig. 4.



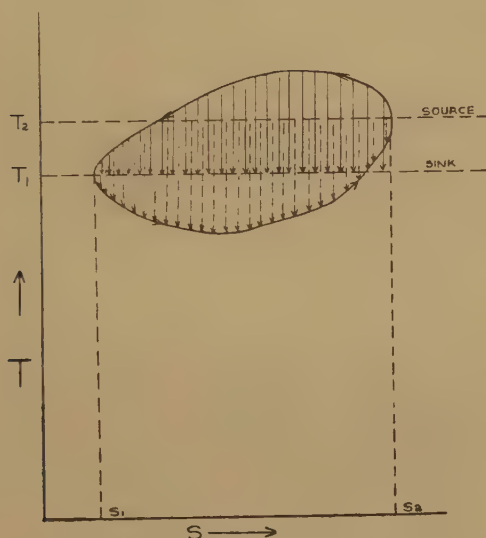
The most interesting feature of the foregoing results is that the comparable Carnot temperature limits are uniquely determinate for every irreversible cycle, whether of the engine or refrigeration type, but that it appears to be possible, in the case of the refrigeration cycle, for the temperature of the Carnot cycle source to be greater than the temperature of the sink. This apparent possible reversal of the relative temperature levels of source and sink in the refrigeration cycle requires, however, to be more closely examined before final conclusions thereon can be drawn. Referring to fig. 5, a hypothetical refrigerating cycle has there been drawn, in which this reversal of source and sink is depicted. As indicated by the arrow-heads showing direction of heat flow, there is nothing inherently impossible in such a cycle, although the refrigerating or cooling action is now being applied to a source of higher temperature than the sink. Clearly the sink could be used directly, by heat conduction or other natural heat flow method, to withdraw heat from the source without the interposition of a refrigerator. The interesting conjecture which



arises here is whether or not it is possible to utilize such a refrigerating cycle for the purpose, either of reducing the bulk of the heat transfer apparatus normally required or of increasing the rate of cooling possible by the more normal direct methods. This point will be given attention later, when considering a particular type of refrigeration cycle.

The writer may, perhaps, be pardoned for introducing the personal note here by stating that, until he had satisfied himself on the several points elucidated above, he has to confess that the problem of settling

Fig. 5.



upon the comparable Carnot temperature limits for some irreversible refrigerating cycles which he had had under analysis, usually baffled him. By the above reasoning, however, such problems are problems no longer.

#### 4. Regenerative Engine Cycles.

The general principles relating to the comparable Carnot temperature limits having been made clear by the foregoing analysis, it is of interest to apply these to any engine cycle in which it is possible to adopt the principle of internal heat regeneration. In a previous communication<sup>(1)</sup> the author has given an analysis of regenerative engine cycles. The results of that analysis led him to make suggestions appertaining to the methods to be followed in the development of internal combustion turbines. The validity of the suggestions there made has since been amply proved by subsequent developments. The purpose of this section is to emphasize more fully the effects of any inefficiency in the heat regenerative apparatus employed. The treatment here given is also necessary as a preliminary to the matters discussed in later sections.

Figs. 6 and 7 are the PV and TS diagrams for the constant pressure regenerative engine cycle.

It is to be noted that the position of point "e" relative to that of point "o" on the entropy diagram is entirely settled by the degree of compression in the cycle, and that it is only when  $T_e > T_o$  that any degree of internal regeneration is possible, as shown by the horizontal arrow-heads in the entropy diagram. In that diagram UT represents the upper and IT the lower temperature limit. Since this analysis is to be used later in application to warming cycles, the lower temperature limit becomes an intermediate temperature in such cycles, hence the use of IT here in place of LT. The thermal efficiency is given generally by

Fig. 6.

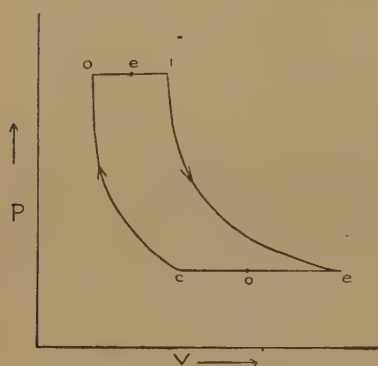
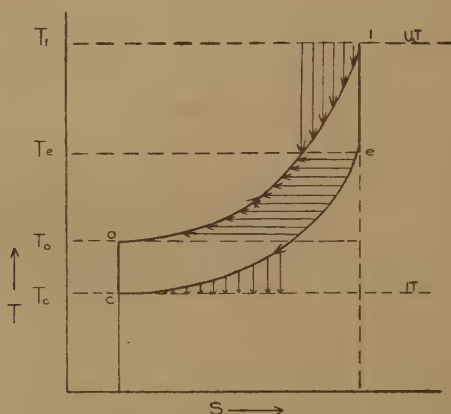


Fig. 7.



(Ext. heat from UT + Int. heat received) —

$$\eta_t = \frac{\text{(Ext. heat to IT + Int. heat rejected)}}{\text{External heat from UT}}$$

$$= \frac{(eH_{EUT} + eH_{IUT}) - (eH_{EIT} + eH_{IIT})}{eH_{EUT}} \quad \text{or}$$

$$\eta_t = \frac{(eH_{EUT} - eH_{EIT}) - (eH_{IIT} - eH_{IUT})}{eH_{EUT}} \quad \dots \dots \dots (1)$$

Since regenerator efficiency is given by

$$\eta_{re} = \frac{eH_{IUT}}{eH_{IIT}}, \quad \dots \dots \dots (2)$$

it follows that for 100 per cent. efficiency of regeneration  $eH_{IUT} = eH_{IIT}$  and the second bracketed expression in (1) will vanish. In practice, however,  $eH_{IIT}$  will always be greater than that of  $eH_{IUT}$ , hence the thermal efficiency will always be less than the ideal value. If we write the ideal efficiency as  $\eta'_t$ , the expression for thermal efficiency becomes

$$\eta = \eta'_t - (1 - \eta_{re}) \frac{eH_{IIT}}{eH_{EUT}} \quad \dots \dots \dots (3)$$

or 
$$\eta_t = \eta_t' - \frac{{}_eH_{IUT}}{{}_eH_{EUT}} \left( \frac{1}{\eta_{re}} - 1 \right) \dots \dots \dots (4)$$

Under these conditions of heat regeneration the actual heat discharged to IT becomes

$$\begin{aligned} {}_eH'_{EIT} &= {}_eH_{EIT} + ({}_eH_{IIT} - {}_eH_{IUT}) \\ &= K_p \{ (T_o - T_c) + (1 - \eta_{re})(T_e - T_o) \} \\ &= K_p \{ (T_e - T_c) - \eta_{re}(T_e - T_o) \}. \dots \dots \dots (5) \end{aligned}$$

These expressions will be utilized later in relation to warming cycles.

### 5. Regenerative Refrigeration Cycles.

Figs. 8 and 9 are the PV and TS diagrams for the constant pressure regenerative refrigeration cycle.

Fig. 8.

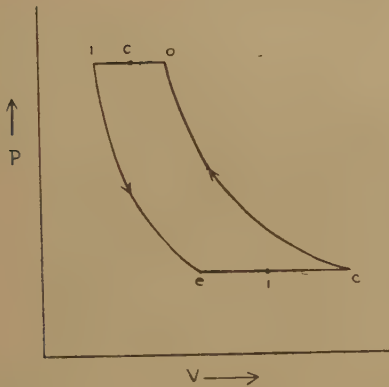
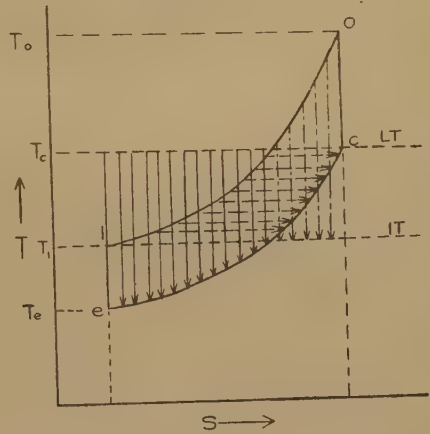


Fig. 9.



The first point to be noted here is that a regenerative refrigerating cycle is impossible unless the temperature of the source, LT, exceeds that of the sink, IT. This results in the apparently anomalous relative positions of LT and IT in fig. 9. The only purpose of such a cycle would be its use for the purpose, if at all feasible, as already stated, of reducing the size of heat transmission apparatus or of increasing the rate of heat transmission. The coefficient of performance is given generally by

$$\begin{aligned} \lambda &= \frac{\text{External Heat from LT}}{(\text{Ext. heat to IT} + \text{Int. heat discharged}) - (\text{Ext. heat from LT} + \text{Ext. heat absorbed})} \\ &= \frac{{}_rH_{ELT}}{({}_rH_{EIT} + {}_rH_{IIT}) - ({}_rH_{ELT} + {}_rH_{ILT})}, \\ \text{or } \lambda &= \frac{{}_rH_{ELT}}{({}_rH_{EIT} - {}_rH_{ELT}) + ({}_rH_{IIT} - {}_rH_{ILT})} \dots \dots \dots (6) \end{aligned}$$

The refrigerator regenerative efficiency is given by

$$\eta_{rr} = \frac{{}_rH_{ILT}}{{}_rH_{ITT}},$$

so that for 100 per cent. efficiency of regeneration,  ${}_rH_{ILT} = {}_rH_{ITT}$ , and the second bracketed expression in the denominator of (6) vanishes. In practice, however,  ${}_rH_{ITT}$  will always be greater than  ${}_rH_{ILT}$ , hence the coefficient of performance will always be less than the ideal value. If we write the ideal coefficient of performance as  $\lambda'$ , the expression for the actual coefficient of performance is obtained by writing

$$\frac{1}{\lambda} = \frac{1}{\lambda'} + \frac{{}_rH_{ITT} - {}_rH_{ILT}}{{}_rH_{ELT}},$$

$$\frac{1}{\lambda} = \frac{1}{\lambda'} + (1 - \eta_{rr}) \frac{{}_rH_{ITT}}{{}_rH_{ELT}},$$

$$\therefore \lambda = \frac{1}{\frac{1}{\lambda'} + (1 - \eta_{rr}) \frac{{}_rH_{ITT}}{{}_rH_{ELT}}} \quad \dots \dots \dots (7)$$

or

$$\lambda = \frac{1}{\frac{1}{\lambda'} + \left( \frac{1}{\eta_{rr}} - 1 \right) \frac{{}_rH_{ILT}}{{}_rH_{ELT}}} \quad \dots \dots \dots (8)$$

Under these conditions of heat regeneration the actual heat discharged to IT becomes

$$\begin{aligned} {}_rH'_{EIT} &= {}_rH_{EIT} + ({}_rH_{ITT} - {}_rH_{ILT}) \\ &= K_p \{ (T_0 - T_c) + (1 - \eta_r)(T_c - T_1) \}, \end{aligned}$$

$$\therefore {}_rH'_{EIT} = K_p \{ (T_0 - T_1) - \eta_r(T_c - T_1) \}. \quad \dots \dots \dots (9)$$

Reference to these expressions will be made when dealing next with warming cycles. Before passing on to these, however, it is interesting to determine the ideal coefficient of performance of a regenerative refrigerator cycle of this type. It is given by

$$\begin{aligned} \lambda &= \frac{{}_rH_{ELT}}{{}_rH_{EIT} - {}_rH_{ELT}} \\ &= \frac{T_1 - T_c}{(T_0 - T_c) - (T_1 - T_c)} \\ &= \frac{1}{\rho - 1}, \quad \dots \dots \dots (10) \end{aligned}$$

$$\text{where } \rho = \frac{T_c}{T_e}.$$



Since, as in the corresponding heat engine cycle,  $\rho$  must be greater than  $r^\gamma - 1$  where  $r = \frac{V_2}{V_1}$ , it follows that for high coefficient of performance  $\rho$  must be small and so, correspondingly,  $r$  must be small. This, as might be expected, is the opposite of the result obtained for the ideal efficiency of the regenerative heat engine cycle, which, as already shown <sup>(1)</sup>, is given by

$$\eta_i = 1 - \frac{1}{\rho},$$

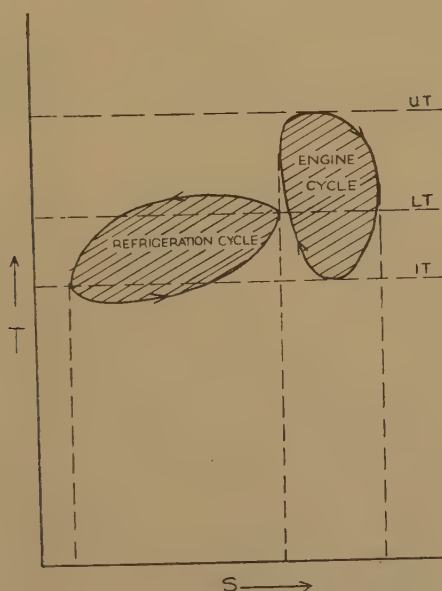
which requires a high value of  $\rho$  for a high efficiency.

The deduction to be drawn from (10) is that in relation to the power required to drive a refrigerator operating under high coefficient of performance conditions on this cycle, the machine will be a bulky one unless high speed rotary compression and expansion is adopted. Nevertheless, it may have advantages for certain purposes.

#### 6. Warming Cycles.

In discussing these, it will be assumed, for the sake of generality as before, that both the engine and the refrigeration cycle are of the

Fig. 10.



regenerative type. Fig. 10 shows the combination of cycles involved. For the case of a non-regenerative refrigeration cycle, IT would be disposed to lie between UT and LT. In any case, it is to be observed

that IT represents the region to be warmed. The efficiency of any warming combination of cycles may be written

$$\begin{aligned}\eta_w &= \frac{\text{Total external heat discharged to IT}}{\text{External heat supplied from UT}} \\ &= \frac{{}_eH'_{\text{EIT}} + {}_rH'_{\text{EIT}}}{{}_eH_{\text{EUT}}} \\ &= \frac{{}_eH_{\text{EIT}} + ({}_eH_{\text{IIT}} - {}_eH_{\text{IUT}}) + {}_rH_{\text{EIT}} + ({}_rH_{\text{IIT}} - {}_rH_{\text{ILT}})}{{}_eH_{\text{EUT}}}. \quad (11)\end{aligned}$$

In addition, we note that

$$\eta_t = \frac{W}{{}_eH_{\text{EUT}}} \quad \dots \dots \dots (12)$$

and

$$\lambda = \frac{{}_rH_{\text{ELT}}}{W}, \quad \dots \dots \dots (13)$$

where  $W$  represents the work done *by* the working fluid in the engine cycle and also the work done *on* that of the refrigeration cycle.

From (1) and (6), however, we see that

$$\begin{aligned}&{}_rH_{\text{EIT}} + ({}_rH_{\text{IIT}} - {}_rH_{\text{ILT}}) \\ &= {}_rH_{\text{ELT}} + ({}_eH_{\text{EUT}} - {}_eH_{\text{EIT}}) - ({}_eH_{\text{IIT}} - {}_eH_{\text{IUT}}),\end{aligned}$$

$\therefore$  (11) becomes

$$\begin{aligned}\eta_w &= \frac{{}_rH_{\text{ELT}} + {}_eH_{\text{EUT}}}{{}_eH_{\text{EUT}}} \\ &= 1 + \frac{{}_rH_{\text{ELT}}}{{}_eH_{\text{EUT}}}, \text{ which, from (12) and (13), gives} \\ \eta_w &= 1 + \lambda\eta_t. \quad \dots \dots \dots (14)\end{aligned}$$

This result, which can be deduced from first principles, for any combination of warming cycles, whether regenerative or non-regenerative, provides a check on the foregoing analysis. The values of  $\lambda$  and  $\eta_t$ , for the general constant pressure cycle, are given by expressions (3), (4), (7) and (8). The chief value of expression (14) lies in the indication it gives of the degree of warming efficiency to be expected in any given combination.

An interesting, and perhaps feasible suggestion, requiring, of course, close technical examination, is to consider the application of the regenerative refrigerating cycle to the cooling of hot mine air by means of water available at a lower temperature. The mine air then becomes the source of heat and the mine water the sink into which the heat is discharged and carried away, not, of course, wholly by conduction, but also by the action of the refrigerating machine and its associated apparatus.

#### Reference.

- (1) "Regeneration and Regenerative Cycles," *Phil. Mag.*, Oct. 1928.

# LXXXIX. *Mathematical Theory of Deflection of Beam.*

By J. G. FREEMAN, M.A., Ph.D. \*†.

[Received August 21, 1945.]

## 1. *Introduction.*

A LIGHT, straight beam is supported on two smooth supports, and a force at right-angles to the beam is applied at a point midway between the supports.

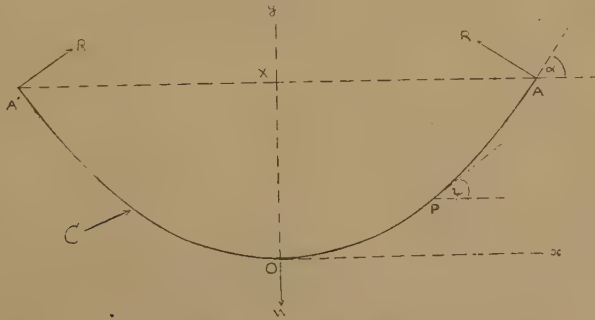
Elementary theory gives the resulting deflection only when this is small compared with the span. Provided, however, that the depth of the beam is small enough, it is possible for the deflection to be comparable with or greater than the span, without the elastic limit being exceeded.

The problem to be treated here is the determination of the relation between the central force, the span and the deflection, when the deflection is not necessarily small compared with the span. Finally, the results obtained are verified experimentally, as described in § 6.

## 2. *Conditions for Equilibrium.*

The beam will be supposed initially to have rectangular cross-section. The surface formed by points of the beam midway between the top and bottom surfaces will be called "the central surface."

Fig. 1.



Curve C, fig. 1, is the section of the central surface by the plane in which lie the lines of action of the reactions R, R at the supports and the force W applied midway between the supports. Lines of action of R, R, W meet C at A, A', O; since the supports are smooth, reactions R, R are normal to C. Ox, Oy are co-ordinate axes respectively parallel and perpendicular to AA'. Span AA' and deflection XO will be denoted by  $2L$  and  $\Delta$ . The tangents to C at P (any point on C between O, A) and A make angles  $\psi$ ,  $\alpha$  with Ox.

\* Central Technical College, Birmingham.

† Communicated by the Author.

By the elementary theory of Strength of Materials there is at P a bending moment  $\frac{IE}{\rho}$ , where I is the second moment of the cross-section through P about the line in which it intersects the central surface, E is the modulus of elasticity of the material considered, and  $\rho$  is the radius of curvature of C at P. Hence,  $x$  and  $y$  being co-ordinates of P,

$$\frac{IE}{\rho} = R \cos \alpha (L-x) + R \sin \alpha (L-y). \quad (2.1)$$

Resolving forces in direction Oy, for the equilibrium of the beam, we have

$$2R \cos \alpha = W. \quad (2.2)$$

### 3. Geometrical Equations.

Differentiation of (2.1) with respect to  $s$ , the arc of C measured from O, gives

$$-\frac{IE}{\rho^2} \frac{d\rho}{ds} = -R \cos \alpha \cos \psi - R \sin \alpha \sin \psi = -R \cos (\alpha - \psi).$$

Since

$$\rho = \frac{ds}{d\psi},$$

$$\frac{IE}{\rho^3} \frac{d\rho}{d\psi} = R \cos (\alpha - \psi). \quad (3.1)$$

Integration with respect to  $\psi$  gives

$$-\frac{IE}{2\rho^2} = -R \sin (\alpha - \psi) + \text{arbitrary constant}.$$

Since, from (2.1),

$$\frac{1}{\rho} = 0$$

at A, when  $\psi = \alpha$ , the arbitrary constant in the above vanishes, and

$$\frac{1}{\rho} = \left\{ \frac{2R}{IE} \sin (\alpha - \psi) \right\}^{\frac{1}{2}}. \quad (3.2)$$

Now

$$\begin{aligned} \frac{dx}{d\psi} &= \frac{dx}{ds} \rho = \left( \frac{IE}{2R} \right)^{\frac{1}{2}} \cos \psi \cdot \sin^{-\frac{1}{2}} (\alpha - \psi), \\ \frac{dy}{d\psi} &= \frac{dy}{ds} \rho = \left( \frac{IE}{2R} \right)^{\frac{1}{2}} \sin \psi \cdot \sin^{-\frac{1}{2}} (\alpha - \psi). \end{aligned}$$

Hence

$$\left. \begin{aligned} \left( \frac{2R}{IE} \right)^{\frac{1}{2}} x &= \int_0^{\psi} \cos \psi \cdot \sin^{-\frac{1}{2}} (\alpha - \psi) d\psi, \\ \left( \frac{2R}{IE} \right)^{\frac{1}{2}} y &= \int_0^{\psi} \sin \psi \cdot \sin^{-\frac{1}{2}} (\alpha - \psi) d\psi. \end{aligned} \right\} \quad (3.3)$$

Writing  $\alpha - \psi = \theta$  and observing that

$$\int_0^{\psi} \cos \theta \sin^{-\frac{1}{2}} \theta d\theta = 2 \sin^{\frac{1}{2}} \psi, \quad (3.4)$$



and putting  $\int_0^\psi \sin^{\frac{1}{2}} \theta \, d\theta = P(\psi), \quad \dots \dots \dots (3.5)$

(3.3) may be written

$$\left. \begin{aligned} \left(\frac{2R}{IE}\right)^{\frac{1}{2}} x &= 2 \cos \alpha (\sin^{\frac{1}{2}} \alpha - \sin^{\frac{1}{2}} \theta) + \sin \alpha \{P(\alpha) - P(\theta)\}, \\ \left(\frac{2R}{IE}\right)^{\frac{1}{2}} y &= 2 \sin \alpha (\sin^{\frac{1}{2}} \alpha - \sin^{\frac{1}{2}} \theta) - \cos \alpha \{P(\alpha) - P(\theta)\}. \end{aligned} \right\} \quad (3.6)$$

In particular, if  $\psi = \alpha$ ,

$$\left. \begin{aligned} \left(\frac{2R}{IE}\right)^{\frac{1}{2}} L &= 2 \cos \alpha \sin^{\frac{1}{2}} \alpha + \sin \alpha \cdot P(\alpha), \\ \left(\frac{2R}{IE}\right)^{\frac{1}{2}} \Delta &= 2 \sin \alpha \sin^{\frac{1}{2}} \alpha - \cos \alpha \cdot P(\alpha). \end{aligned} \right\} \quad \dots \dots (3.7)$$

The elimination of  $\alpha$  between these equations, after the substitution  $2R = W \sec \alpha$ , gives the required relation between  $W$ ,  $\Delta$ ,  $L$  \*.

For brevity I shall write

$$\left. \begin{aligned} 2 \cos \alpha \cdot \sin^{\frac{1}{2}} \alpha + \sin \alpha \cdot P(\alpha) &= Q(\alpha), \\ 2 \sin \alpha \cdot \sin^{\frac{1}{2}} \alpha - \cos \alpha \cdot P(\alpha) &= S(\alpha), \end{aligned} \right\} \quad \dots \dots (3.8)$$

so that (3.7) may be written

$$\left(\frac{2R}{IE}\right)^{\frac{1}{2}} L = Q(\alpha), \quad \left(\frac{2R}{IE}\right)^{\frac{1}{2}} \Delta = S(\alpha), \quad \dots \dots \dots (3.9)$$

whence

$$\frac{\Delta}{L} = \frac{S(\alpha)}{Q(\alpha)}, \quad \dots \dots \dots (3.10)$$

and, from (3.9) and (2.2),

$$\left(\frac{W}{IE \cos \alpha}\right)^{\frac{1}{2}} L = Q(\alpha),$$

whence

$$\frac{WL^2}{IE} = Q(\alpha)^2 \cos \alpha, \quad \dots \dots \dots (3.11)$$

and, from (3.6), (3.7) and (3.8),

$$\left. \begin{aligned} \frac{Q(\alpha)}{L} x &= Q(\alpha) - 2 \cos \alpha \sin^{\frac{1}{2}} \theta - \sin \alpha \cdot P(\theta), \\ \frac{S(\alpha)}{\Delta} y &= S(\alpha) - 2 \sin \alpha \sin^{\frac{1}{2}} \theta + \cos \alpha \cdot P(\theta). \end{aligned} \right\} \quad \dots \dots (3.12)$$

These are the equations to the deflection curve in terms of parameter  $\theta$ .

---

\* This cannot be done algebraically, but a graphical solution is given in § 4

4. *Numerical Calculations.*

Reading values of  $\sin^{\frac{1}{2}}\theta$  at intervals of  $2\frac{1}{2}^\circ$ , and evaluating P at intervals of  $5^\circ$  by Simpson's Rule, the following results are obtained, correct to two decimal places \* :—

$\alpha$ degrees.	P.	Q.	S.	$S/Q$ $=\Delta/L$ .	$Q^2 \cos \alpha$ $=WL^2/IE$ .
0	0.00	0.00	0.00	0.00	0.00
5	0.02	0.59	0.03	0.06	0.35
10	0.05	0.83	0.10	0.12	0.68
15	0.09	1.01	0.18	0.18	0.99
20	0.14	1.15	0.27	0.24	1.24
25	0.19	1.26	0.38	0.30	1.44
30	0.25	1.35	0.49	0.36	1.58
35	0.31	1.42	0.61	0.43	1.65
40	0.38	1.47	0.74	0.50	1.67
45	0.45	1.51	0.87	0.57	1.61
50	0.53	1.53	1.00	0.65	1.51
55	0.61	1.53	1.14	0.74	1.36
60	0.69	1.53	1.27	0.84	1.17
65	0.77	1.50	1.40	0.94	0.95
70	0.85	1.46	1.54	1.05	0.73
75	0.93	1.41	1.66	1.18	0.52
80	1.02	1.35	1.78	1.32	0.32
85	1.10	1.27	1.88	1.48	0.14
90	1.19	1.19	2.00	1.68	0.00
95	1.28	1.10	2.09	1.90	-0.11
100	1.37	1.00	2.19	2.20	-0.17
105	1.45	0.89	2.28	2.55	-0.21
110	1.54	0.78	2.35	3.00	-0.21
115	1.62	0.66	2.41	3.66	-0.18
120	1.70	0.54	2.46	4.59	-0.14
125	1.78	0.42	2.50	6.03	-0.10
130	1.86	0.30	2.53	8.60	-0.06
135	1.93	0.18	2.55	14.50	-0.02
140	2.00	0.06	2.57	44.20	-0.00
145	2.07	-0.05	2.56	-51.30	-0.00
150	2.13	-0.16	2.55	-16.08	-0.02
155	2.19	-0.25	2.53	-10.02	-0.06
160	2.25	-0.33	2.51	-7.63	-0.10
165	2.29	-0.39	2.48	-6.33	-0.15
170	2.34	-0.41	2.45	-5.91	-0.17
175	2.37	-0.38	2.41	-6.30	-0.15
180	2.38	-0.00	2.38	$-\infty$	-0.00

\* Alternatively,  $P(\alpha)$  may be expressed as an elliptic integral

$$2 \int_{\frac{\pi}{4} - \frac{\alpha}{2}}^{\frac{\pi}{4}} (1 - 2 \sin^2 u)^{\frac{1}{2}} du$$

values of which are given in tables.

From this table \* are plotted graphs :—

(A)  $\alpha$  against  $\Delta/L$ ,

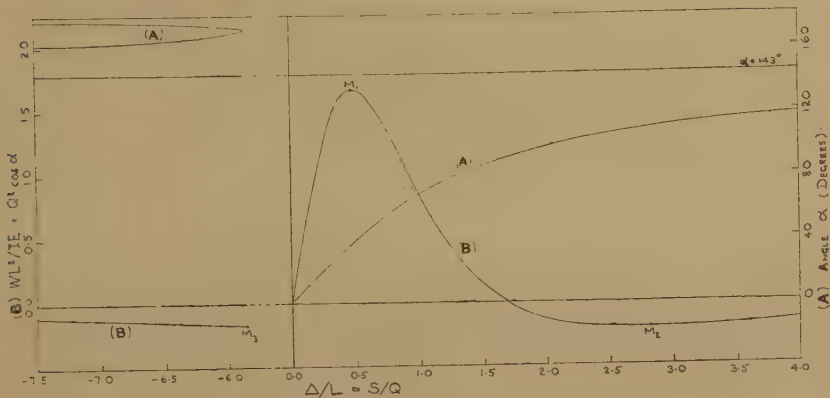
(B)  $WL^2/IE$  against  $\Delta/L$ .

These are shown in fig. 2.

Plotting the graph of  $Q$  against  $\alpha$  shows that  $Q$  is positive for values of  $\alpha$  between  $0^\circ$  and  $143^\circ$  (approximately), and negative for values of  $\alpha$  between  $143^\circ$  and  $180^\circ$ . In the former case  $L$  is positive, in the latter negative, and the beam crosses over itself (fig. 3).

When the beam crosses over itself, the problem is no longer two-dimensional, but, provided the beam is narrow, the theory given here will be approximately true.

Fig. 2.



The following problems may now be solved graphically :—

- (1) From measurements of  $L$ ,  $\Delta$ ,  $W$ ,  $I$  to determine  $E$ .
- (2) Given  $L$ ,  $W$ ,  $I$ ,  $E$  to determine  $\Delta$ ,  $R$ ,  $\alpha$ .
- (3) Given  $\Delta$ ,  $L$  to determine the deflection curve.

Method :—

(1) Evaluate  $\Delta/L$ . Read  $WL^2/IE$  from graph (B).  $W$ ,  $L$  and  $I$  being known,  $E$  is then determined.

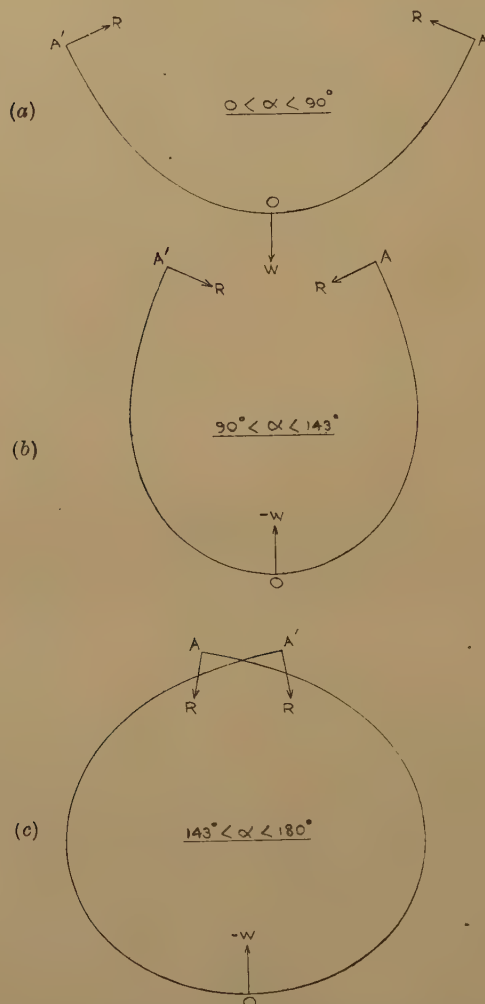
(2) Evaluate  $WL^2/IE$ . Read possible values of  $\Delta/L$  from graph (B);  $L$  being known,  $\Delta$  is then determined. Read  $\alpha$  from graph (A);  $R$  is determined by  $R=W/2 \cos \alpha$ .

\* Together with additional values :—

$\alpha$ .	$S/Q$ .	$Q^2 \cos \alpha$ .
171	-5.85	0.17
172	-5.90	0.17
176	-6.68	0.13
177	-7.21	0.11

(3) Evaluate  $\Delta/L$ . Read  $\alpha$  and  $Q^2 \cos \alpha$  from graphs (A), (B);  $Q$  is thus determined; determine  $S$  from  $S/Q = \Delta/L$ . Giving  $\theta$  values  $0, 5^\circ$ ,

Fig. 3.



$10^\circ, 15^\circ \dots$  (not exceeding  $\alpha$ ), and reading values of  $P(\theta)$  from table of this paragraph,  $x, y$  are given by (3.12), enabling curve  $OA$  to be plotted; curve  $OA'$  is obtained by reflection of  $OA$  in  $Oy$ .

### 5. Properties of Beam with Constant Span.

(i)  $0 < \alpha < 143^\circ$ .—Beam not crossing over itself.

From graph (B),  $WL^2/IE$  has maximum value 1.66 (at  $M_1$ ) when  $\Delta/L = 0.48$ , and minimum value  $-0.22$  (at  $M_2$ ) when  $\Delta/L = 2.80$ . If  $L$  is constant, the beam thus has the properties shown in fig. 4.

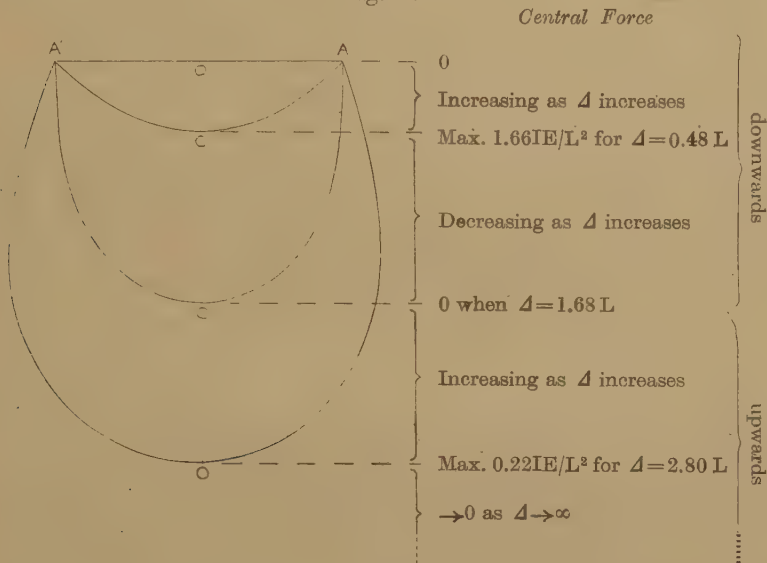


(ii)  $143^\circ < \alpha < 180^\circ$ .—Beam crossing over itself.

From graph (A) it is seen that there are two values of  $\alpha$  for a given value of  $\Delta/L$ ; there are thus two possible deflection curves for a given deflection. Also  $-\Delta/L$  has a minimum value 5.85; equilibrium is therefore not possible with deflection less than  $-5.85 L$ .

From graph (B),  $WL^2/IE$  has minimum value  $-0.17$  (at  $M_3$ ), so that maximum value of central force is  $0.17 L^2/IE$ .

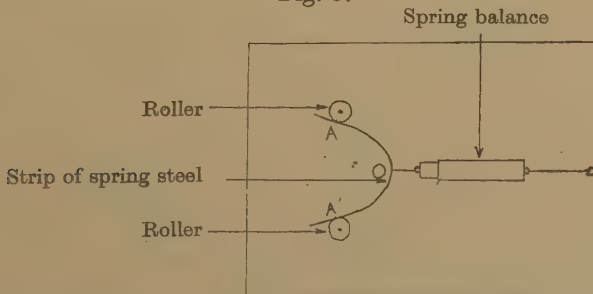
Fig. 4.



### 6. Experimental Verification.

A strip of spring steel, 1 in.  $\times$  0.01 in., was mounted on a horizontal board, pressing against adjustable rollers at A, A' (ensuring normal reactions at these points), and the force at O was supplied and measured by a spring balance (see fig. 5). The strip was supported by glass rods resting on the

Fig. 5.

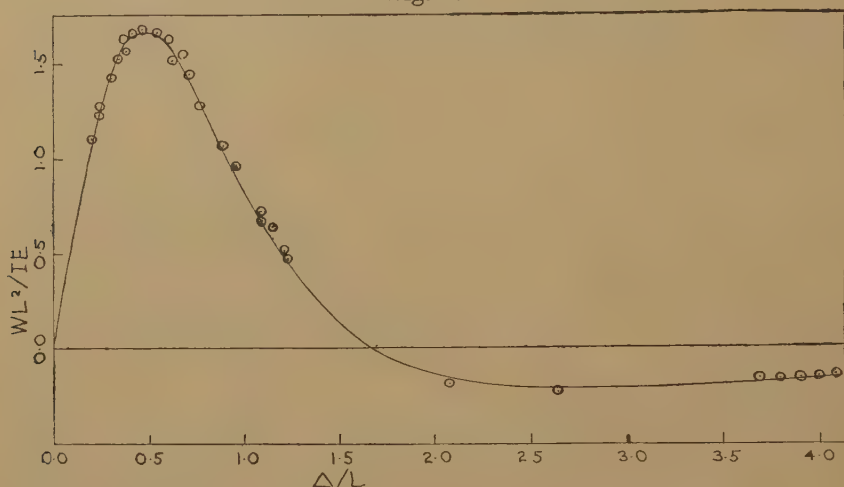


board; these, being free to roll, supported the strip with negligible friction.

By adjusting the positions of the rollers and spring balance, various spans and deflections were obtained, and, taking  $E$  for spring steel to be

$31 \times 10^6$  lb./in.<sup>2</sup>, corresponding values of  $WL^2/IE$  and  $\Delta/L$  were evaluated and compared with the theoretical graph. Results are shown in fig. 6.

Fig. 6.



Continuous curve obtained from theoretical tables.

Points  $\circ$  obtained by experiment.

Secondly, the actual curve formed by the strip for given values of  $\Delta$ ,  $L$  was compared with the theoretical curve with the same values of  $\Delta$ ,  $L$ , as obtained by method given in problem (3) of § 4. The result is shown in fig. 7.

Fig. 7.



Continuous curve is deflection curve obtained by experiment with  
 $\Delta=3$  in.,  $L=1$  in.

Points  $\circ$  are obtained by theory, for  $\theta=0^\circ, 10^\circ, 20^\circ, 30^\circ, \dots, \alpha$  when  
 $\Delta=3$  in.,  $L=1$  in.

It will be seen that there is close agreement between experiment and the theory developed here.

In conclusion, I should like to express my thanks to Messrs. Bowling, Barmby and Densfield, of the Birmingham Central Technical College, for encouragement and assistance given in the preparation of this paper.

XC. *A Laboratory Investigation of Electrosmosis in Soils*\*.

By D. J. MACLEAN, B.Sc., A.Inst.P., and D. W. ROLFE,  
of the Road Research Laboratory,  
Department of Scientific and Industrial Research †.

[Received April 24, 1945.]

## SUMMARY.

A study has been made of the phenomenon known as electrosmosis in which, when an electric current is passed through a mass of wet soil, the soil is dried out around the anode and water is expelled from the cathode. The effect was investigated by means of a simple laboratory method in which a small electric current was passed through a cylindrical specimen of wet soil, the current density being kept low enough to eliminate appreciable errors due to Joule heating.

Moisture movements were found to be greater in sandy than in clay soils. For a given soil the amount of water expelled at the cathode was proportional to the quantity of electricity passed through the specimen, up to the point when the soil at the anode became too dry to act as a conductor. On the assumption that the moisture movement is due to ions carrying water molecules from the anode to the cathode, calculations showed in the case of one of the soils that the number of water molecules transported per ion was 88.

*Introduction.*

WHEN two electrodes are inserted in wet soil and a direct current is passed between them, water moves from the positive electrode to the negative. This effect is one of the electrokinetic phenomena and is called electrosmosis.

Electrosmosis has a potential application to soil drainage and has, in fact, been used in a few civil engineering projects in Germany\*. The simple laboratory experiments described in this paper were undertaken as part of an exploratory investigation at the Road Research Laboratory, and with the aim of obtaining an understanding of the physics of electrosmosis as applied to this particular field.

*Theoretical Explanation of Electrosmosis.*

Before describing the experimental method used in the present investigations, it is perhaps useful to summarize briefly two explanations that have been offered of the electrosmosis effect. Both are, in fact,

---

\* Crown copyright reserved.

† Communicated by the Authors.

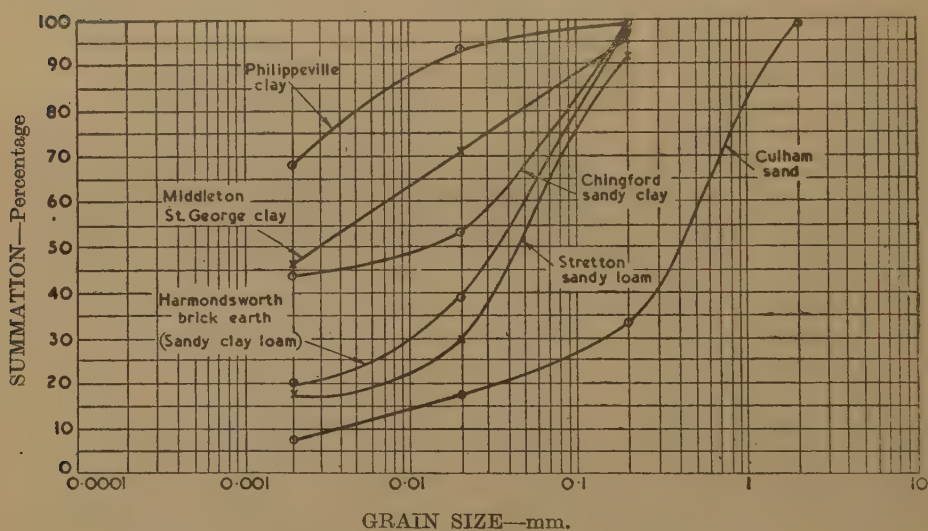
\* Casagrande, L., "Electrical Drainage of Fine-grained Soils," *Strasse*, viii. (19/20), pp. 324-6 (1941).

based on the same electrokinetic theory, but the first probably gives a better visual impression of the phenomenon:—

(1) Positive ions attached to the clay particles are liberated and subsequently migrate to the cathode under the influence of the electric field. Each ion acts as a nucleus to a number of molecules of water. When the ion reaches the negative electrode it gives up its charge and deposits the water which it has carried with it.

(2) The second explanation (N. K. Adam, "Physics and Chemistry of Surfaces, 1941" (3rd Ed.), Chapter 8, p. 351) is based on the fact that an "electric double layer" is set up at almost any boundary between two phases of matter, which results in a difference of potential being

Fig. 1.



Grain-size curves of soils used in electrosmosis tests.

set up between any two phases in contact. In the case of wet soil, the water phase will be positively charged and the soil particles negatively charged. When an electric field is applied to the wet soil, the soil and water tend to move in opposite directions, but on account of the immobility of the soil particles only the water moves.

### *Experimental Method.*

After the preliminary experiments described in the next section the following experimental method was adopted throughout the investigation.

The soil was dried, broken down to a powder and remixed with water to form a mortar of uniform consistency, which was then compacted to a given bulk density and moisture content in a cylindrical mould. Most of the experiments were carried out on cylinders 4 in. in diameter and



4.6 in. long, and the initial moisture content of the soil was normally between its plastic and liquid limits\*.

After ejection from the mould electrodes were formed on the ends of the soil cylinder by coating each surface with graphite, followed by a layer of aluminium foil. A thick coating of shellac was applied to the curved face of the cylinder to prevent evaporation losses and the specimen was then weighed. To eliminate possible errors due to gravitational movements of moisture the specimen was supported for the test in a horizontal position.

A direct current was then passed through the specimen, and intermittent recordings of the loss of moisture at the cathode end were made by unclamping the cylinder, drying the cathode end with cotton wool and weighing. Variations in the current strength and potential difference across the specimen were noted throughout the test.

In some cases at the conclusion of the test the moisture gradient through the specimen between the anode and the cathode was determined immediately after switching off the current. This was done by the ordinary method of sampling.

#### *Preliminary Experiments.*

The local (Harmondsworth) brickearth at a moisture content of 17 per cent. was used in the initial experiments, in which the effect of current strength was investigated. At first the aluminium electrodes were clamped directly on to the ends of the soil cylinder.

It was found that considerable heating due to the Joule effect (resistance heating) occurred, partly in the body of the soil but mainly at the anode end, where the contact resistance between the electrode and the soil increased rapidly as the soil became drier. Fig. 2 shows a continuous record of the temperatures of the soil near the anode, in the middle of the specimen and near the cathode when a current of 150 milliamperes was used. It will be seen that a temperature of 46° C. was reached near the anode, and that the maximum temperatures at the middle of the specimen and at the cathode were reached some time later, indicating that most of the heat was generated at the anode.

It was found that Joule heating increased the movement of moisture through the soil in three ways:—(1) the existence of a temperature gradient caused moisture to move towards the cooler parts of the soil; (2) the increase in temperature lowered the viscosity of the soil water and therefore increased the relative permeability of the soil to moisture movement; (3) heat tended to increase evaporation losses. In the case of a cylinder of brickearth, 8 gm. of water was expelled by 1,000 coulombs of electricity when an initial current of 70 milliamperes was used, while a similar quantity of electricity expelled 20 gm. of water when the initial

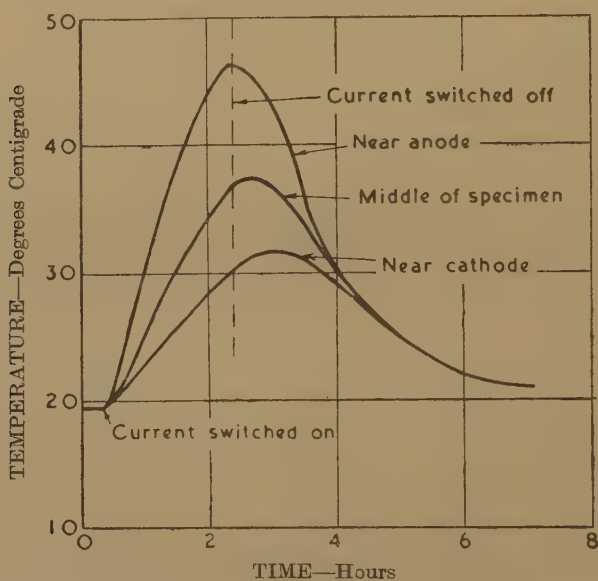
---

\* The plastic limit of a soil is the moisture content below which the soil is friable and above which it is plastic; the liquid limit is the moisture content below which the soil is plastic and above which it is fluid. These limits are measured by standard methods.

current was 150 milliamperes. The water movement produced by Joule heating, therefore, was greater than that due to the electrosmosis effect alone, since in the conditions of the test, as shown later, the electrosmosis effect is proportional to the quantity of electricity used.

As the purpose of the investigation as a whole was to study the electrosmosis effect, steps had to be taken to reduce the subsidiary effect to negligible proportions. It was found that this contact heating was largely eliminated when the ends of the specimen were coated with graphite, and that by limiting the current to 70 milliamperes the rise of temperature could be limited to 2 or 3° C. This procedure was adopted in all subsequent tests.

Fig. 2.



Temperature record showing the effect of Joule heating due to high contact resistance at anode.

### *The Relation between Water Movement, Electrical Units and Shape of Soil Specimen.*

A series of specimens was tested to determine the respective parts played by the quantity of electricity (in coulombs) and the electrical energy (in watt-hours) expended in driving the current through the soil, and to study the effect of the shape and volume of the specimen on the amount of water expelled.

Four cylinders of brickearth were therefore prepared at the same initial moisture content and dry density (*i. e.*, bulk density of solids in soil), each cylinder having either a different length or diameter. The four cylinders were connected in series to a source of electrical supply so that each specimen had the same quantity of electricity passed through

it, but the electrical energy consumed was different in each case because the resistance of each specimen was different. The amounts of water expelled after the passage of 1,000 coulombs of electricity are reported in Table I., together with the values of the resistance and potential difference across the ends of each specimen and the number of watt-hours expended in passing this quantity of electricity through the specimens.

Within the limits of experimental error the amount of water expelled from each specimen was the same. The amount of water moved thus depends on the quantity of electricity passed through the soil. The effect of the shape and volume of the soil is to vary the resistance between the electrodes and, in consequence, to vary the number of watt/hours required to drive a given quantity of electricity through the soil and to expel a given quantity of water.

TABLE I.

The relation between the amount of expelled water, electrical energy consumed, and shape of soil specimen when 1,000 coulombs had been passed through each specimen.

Dimensions of cylindrical specimens	Weight of expelled water (gm.)	Potential difference across specimens (volts)	Current through specimens (amperes)	Resistance of specimens (ohms)	Energy consumed (watt-hours)	Quantity of electricity passed (coulombs)
5" high × 3¼" dia.	7.7	48	0.064	750	10.2	1,000
4" high × 3¼" dia.	8.0	34	0.064	530	8.0	1,000
3" high × 3¼" dia.	7.2	38	0.064	595	7.3	1,000
2.6" high × 4" dia.	8.4	14	0.064	220	3.9	1,000

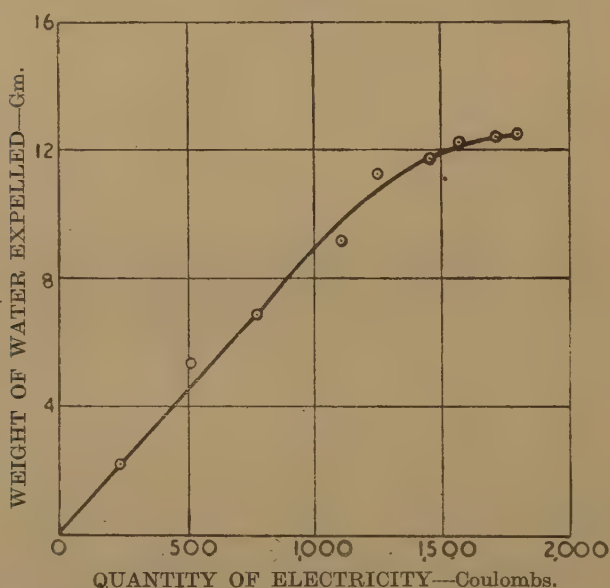
*The Relation between Quantity of Water Expelled and the Quantity of Electricity Used.*

The relation between the quantity of electricity used and the quantity of water expelled is shown graphically in fig. 3. The curve, which was obtained for brickearth, is typical of those obtained for other types of soil. Up to 1,000 coulombs a linear relation exists between the amount of water expelled and the quantity of electricity. After this the quantity of water expelled gradually decreases until a further passage of current expels no more water. The departure from linearity is associated with a rapid increase in the resistance of the specimen, by virtue of which the current decreases to an extremely low value.

*Comparison of the Electrosmosis Effect in Different Types of Soils.*

Soils covering a wide variety of types were made into specimens having the same degree of packing and the same moisture content as those referred to in Table II. The weights of water expelled from these cylinders, corresponding to various quantities of electricity passed through them, were determined and plotted in a similar manner to fig. 3. In all cases the departure from linearity occurred for quantities of electricity greater than 1,000 coulombs. The weights of water expelled per 1,000 coulombs are given in Table II., which also gives details of the soils used. The grain-size analyses of the soil are compared in fig. 1.

Fig. 3.



Relation between weight of water expelled and the quantity of electricity used for Harmondsworth brickearth.

The weight of water expelled per 1,000 coulombs was greatest for sandy soils and least for clays. The number of coulombs required to expel 1 gm. of water from the various soil cylinders has been plotted in fig. 4 against the "clay" content of each soil (the "clay" grade is defined as the material finer than 0.002 mm.). It will be seen that the points lie close to a straight line.

*The Effect of the Bulk Density of the Soil on the Amount of Water Expelled.*

In order to determine the effect of the bulk density of the soil on the amount of water expelled, three cylinders of soil were made up at different densities but the same moisture content from both Harmondsworth brickearth and Culham sand. Electric current was passed through the

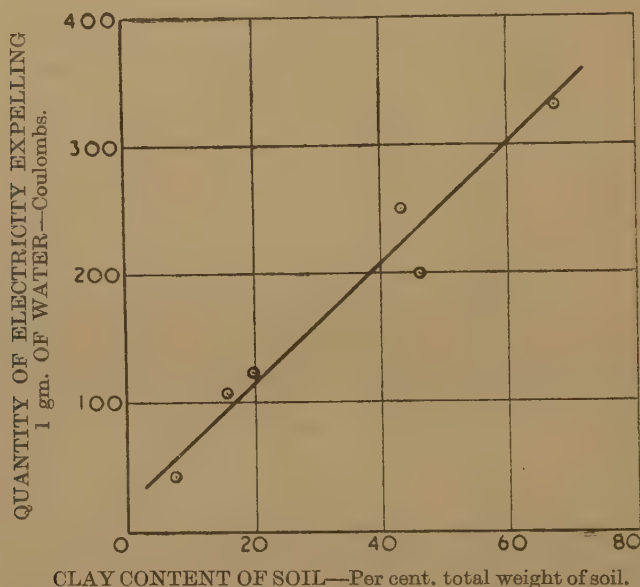


TABLE II.  
Weight of water expelled per 1,000 coulombs for various types of soil.

Location of soil.	Soil type.	Liquid limit. (%)	Plastic limit. (%)	Plasticity index.	Mechanical analysis (International system)				Initial moisture content. (%)	Weight of water expelled (gm. per 1,000 coulombs)
					Coarse sand (2-0.2 mm.)	Fine sand (0.2-0.02 mm.)	'Silt' (0.02-0.002 mm.)	Clay (<0.002 mm.)		
Culham . . . . .	Sand	Non-plastic			66	17	10	7	18	30
Stretton . . . . .	Sandy loam	26	18	8	8	62	13	17	19	9
Harmondsworth	Sand clay loam	30	21	9	3	59	18	20	17	8
Chingford . . . . .	Sandy clay	54	25	29	0	47	10	43	24	4
Middleton St. George. . . . .	Clay	61	33	28	3	24	26	47	23	5
Philippeville. . . . .	Clay	67	31	36	0	6	25	69	30	3

three cylinders connected in series, and the loss of water from the cathode determined at various times. For both soils the weight of water expelled per coulomb from the cathode was the same at the two higher dry densities (106 and 112 lb./cu. ft. for brickearth, and 114 and 120 lb./cu. ft. for Culham sand), but was lower and variable at the lower dry densities (101 lb./cu. ft. for brickearth and 108 lb./cu. ft. for Culham sand). This is probably because the soil water system was discontinuous at the lower soil densities.

Fig. 4.



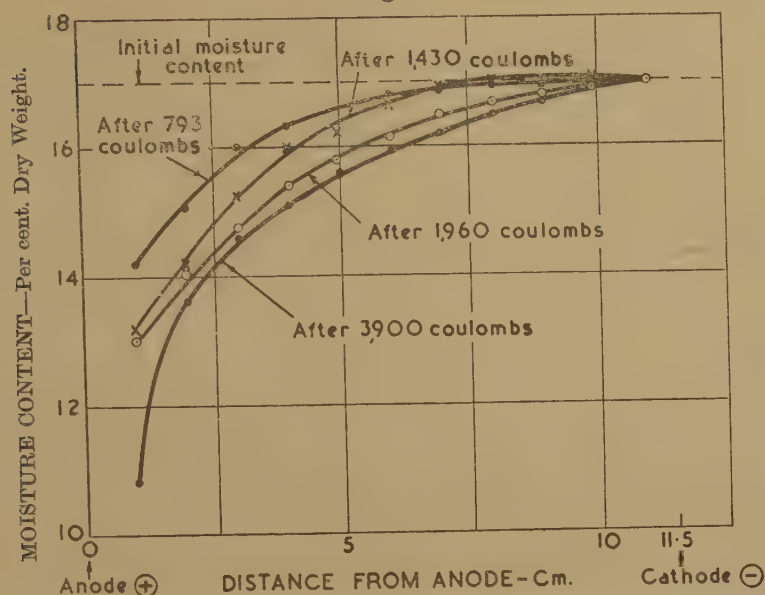
Relation between quantity of electricity required to expel 1 gm. of water from a soil and the clay content of the soil.

#### *Nature of Moisture Movement in Soil.*

Moisture gradients through brickearth cylinders, which were initially at the same uniform moisture content, were determined after varying quantities of electricity had been passed through them. These moisture gradients are shown graphically in fig. 5. Over the range of the experiments, increasing quantities of electricity produced no decrease in moisture content at the cathode and the amount of drying out increased in an exponential manner to a maximum at the anode.

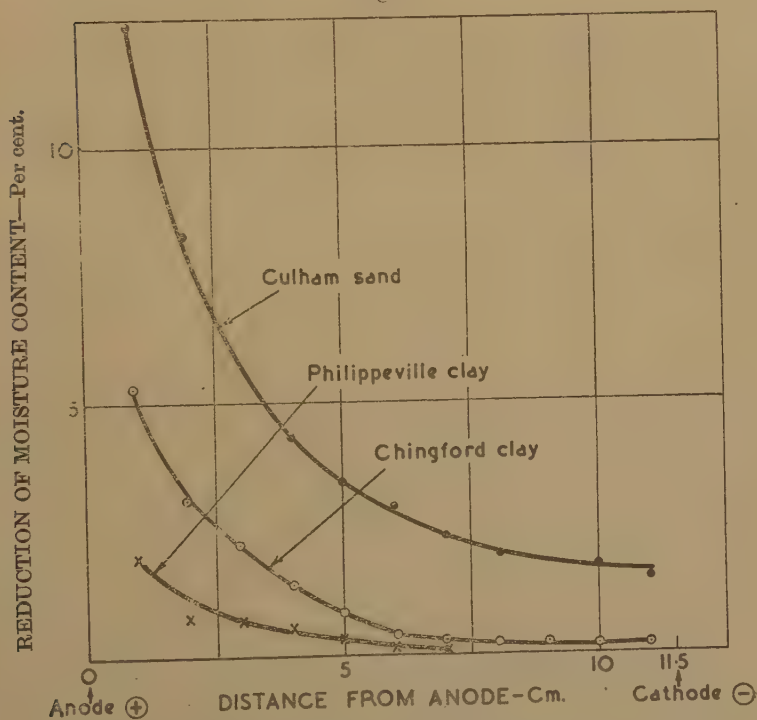
The moisture gradients in cylinders of Culham sand, Chingford clay and Philippeville clay were determined when they reached a condition where no more water could be expelled. The initial moisture contents of the soils were chosen so that each soil was initially in a similar condition of "wetness." In fig. 6, the reduction in moisture content from these original values to their ultimate minimum values have been plotted

Fig. 5.



Moisture gradients in brickearth cylinders after varying quantities of electricity had been passed through them.

Fig. 6.



Limiting moisture gradients in various soil cylinders when no more water can be expelled.

against distance from the anode. The curves confirm the previously-stated conclusion that sandy soils can be dried out to a much greater extent than clay soils.

*Abnormal Effect in Harmondsworth Brickearth.*

In a few cases an abnormal effect was observed with Harmondsworth brickearth in that exactly twice the normal amount of water was expelled per coulomb. This effect was not observed in any of the other soils, and occurred only twice out of twenty or so cylinders of brickearth. Apart from its rarity of occurrence the abnormality of the effect is made clear when the weights of water expelled per 1,000 coulombs under the normal and abnormal conditions are compared with those for other soils.

In terms of ionic motion, the abnormality may be explained by water molecules being carried by ions having a single instead of a double charge, in which case double the number of ions per coulomb would reach the cathode. This assumes that the number of molecules of water carried by each ion is independent of the ionic charge.

*Calculation of the Number of Water Molecules Transported per Ion.*

The electrical resistance of a dried-out cylinder of brickearth as used in the laboratory tests was found to be about 50 megohms. To a close degree of approximation, therefore, it may be taken that the electric current in wet soil flows almost entirely through the water and not through the soil particles.

The number of molecules of water in M gm. is given by:—

$$n = \frac{M}{m},$$

where  $m$  = mass of one molecule of water (gm.).

The number of ions equivalent to a known quantity of electricity may be determined from the equation:

$$Q = Nxe \text{ or } N = \frac{Q}{xe},$$

where  $Q$  = quantity of electricity (e.m.u.),

$N$  = number of ions,

$e$  = electronic charge (e.m.u.),

$x$  = number of electronic charges on each ion.

Therefore, the number of molecules of water transported per ion  $y$  is given by

$$y = \frac{n}{N} = \frac{M}{m} \frac{xe}{Q}.$$

As an example, 8 gm. of water per 1,000 coulombs were normally expelled from brickearth. As brickearth contains a fair proportion of calcium salts the ionic charge will be twice the electronic charge. As



stated above, an abnormal case, in which 16 gm. of water per 1,000 coulombs were expelled, was also observed corresponding to a single-charged ion. As the product  $Mx$  remains constant, both cases give the same value of  $y$ . For the normal effect in brickearth:—

$$M=8 \text{ gm.}, x=2, Q=100 \text{ e.m.u. } (=1,000 \text{ coulombs}),$$

$$m=2.99 \times 10^{-23} \text{ gm.}, e=1.59 \times 10^{-20} \text{ e.m.u.}$$

Hence, 
$$y = \frac{8 \times 10^{23}}{2.99} \times \frac{2 \times 1.59}{100 \times 10^{20}} = 88.$$

### Conclusions.

(1) On passing an electric current through wet soil, the soil is dried out at the anode and water is expelled from the soil at the cathode.

(2) When large currents are employed, considerable heating takes place at the anode. The temperature gradients thus produced increase the water movement caused by the electrosmosis effect alone.

(3) The weight of water expelled at the cathode, due to electrosmosis alone, is proportional to the quantity of electricity in coulombs passed through the soil until the soil dries out at the anode, after which the rate of expulsion of water rapidly decreases to zero.

(4) The weight of water expelled by a given quantity of electricity is highest for sandy soils and lowest for heavy clay soils. A linear relation exists between the quantity of electricity required to expel 1 gm. of water and the "clay" content of the soil.

(5) The weight of water expelled per coulomb is independent of the bulk density of the soil when well compacted. When the soil is loosely compacted the weight lost per coulomb becomes more variable and lower, probably due to a breakdown in the continuity of the soil water system.

(6) The drying out produced by electrical means occurs mainly at and near the positive electrode. The moisture content of the soil at the negative electrode is not changed to any great extent.

### Acknowledgment.

The experimental work described in this paper was carried out at the Road Research Laboratory as part of the programme of the Road Research Board. The paper is published by permission of the Director of Road Research.

Road Research Laboratory,  
Harmondsworth,  
West Drayton,  
Middlesex.

XCI. *On the Raman Spectrum of *p*-Diphenyl Benzene.*

By S. K. MUKERJI, M.Sc., Ph.D.(Lond.), Professor of Physics, Agra College, Agra, and LAKSMAN SINGH, M.Sc., Agra University Research Scholar in Physics \*.

[Received July 19, 1945.]

[Plate XV.]

## ABSTRACT.

The Raman Spectrum of *p*-diphenyl benzene in the solid state and also in the molten state has been investigated for the first time. The spectrum of this substance resembles that of diphenyl and also of the ortho and the meta compound of this series investigated by one of the authors and S. Abdul Aziz <sup>(1)</sup>. The fluorescence observed was very greatly marked in this case. The substance in the solid state, nevertheless, yielded 24 new lines not recorded before. In the molten state only about a dozen lines were observed and recorded, for the first time, owing to very intense fluorescence and experimental difficulties.

The frequencies of these lines in the solid state are at 3041 (3), 2824 (0), 2205 ( $\frac{1}{2}$ ), 2140 ( $\frac{1}{2}$ ), 2095 (3), 2029 (3), 1760 (1), 1673 (0), 1605 (10), 1592 (10), 1549 (3), 1503 (1), 1372 (0), 1274 (10), 1219 (6), 1148 (0), 1039 (1), 1008 (4), 981 (4), 811 (4), 773 (5), 598 ( $\frac{1}{2}$ ), 85 (5) and 42 (5)  $\text{cm}^{-1}$  respectively.

The frequencies observed in the molten state are at 3043 (2), 1600 (10), 1523 (2 diff.), 1359 (1), 1283 (10), 1223 (4), 1007 (5), 940 (2), 761 (2), 593 (0), 409 (1), and 93 (3) ?  $\text{cm}^{-1}$  respectively.

The results have been discussed with reference to the Raman spectra of diphenyl, of ortho- and meta-diphenyl benzene.

A comparative study of the spectra observed in diphenyl and *p*-diphenyl benzene lends support to the view that the molecule of the latter compound is planar in form.

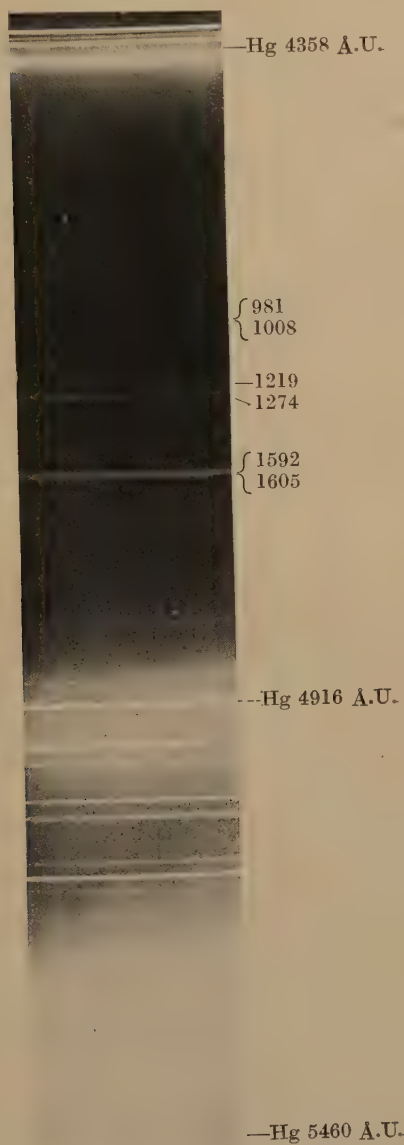
*Introduction.*

THERE appears to have been no investigation done so far on the Raman spectrum of *p*-diphenyl benzene of the diphenyl benzene family. In two previous communications <sup>(2)</sup> one of us and S. Abdul Aziz have submitted the results of our investigations on the ortho and the meta compound of this series. In this investigation we have obtained the Raman spectra of the para compound in the solid and the molten state. In spite of the very strong fluorescence shown on the plates the substance has yielded 24 new lines in the solid state and 12 new lines in the molten state, not recorded before. These results have been discussed here with reference to the Raman lines of the ortho and the meta compound and also of diphenyl.

---

\* Communicated by Prof. S. K. Mukerji.

(Excitation by Hg 4358 Å. only)



*p*-diphenyl benzene (solid state).





As regards the structure of *p*-diphenyl benzene the only work which so far appears to have been done is that due to L. W. Pickett<sup>(3)</sup>, who has made an X-ray study of the compound and by the method of trial and error and Fourier analysis has come to the conclusion that the molecule of this compound is planar in form. As a result of this investigation it has been possible to compare the Raman spectra of *p*-diphenyl benzene with that of diphenyl. The molecule of *p*-diphenyl benzene may be regarded as made of two diphenyl molecules having the central benzene ring common to its two end rings. The spectra of these two compounds have shown some remarkable similarity as regards the frequencies of some prominent lines, lending support to the view that the molecule of *p*-diphenyl benzene is planar in form.

Another interesting comparison now made possible is with the spectra of the ortho, meta and the para compounds of the diphenyl benzene family in so far as the effect of the successive loading of the carbon atoms in the ortho, meta and para positions by a phenyl group is concerned.

#### *Experimental.*

*p*-diphenyl benzene obtained from the research laboratory of Eastman Kodak Company was further purified by crystallization from absolute alcohol. The pure crystals thus formed were then used for obtaining their spectra. The melting point of this compound is 209–210°C., which is very inconveniently high for use in the molten state. In the molten state the substance very often got charred during the course of the exposure, which precluded considerably the chances of getting any good and reliable photographs. The substance was therefore first tried in the solid state with the purest crystals that could be formed by repeated crystallization. The result was very satisfactory. With mercury arc spectrum as the exciting light very intense scattering was obtained and, in spite of very powerful fluorescence, as many as 24 new Raman lines were obtained. The fluorescence was, of course, considerably reduced by using a filter of a concentrated solution of sodium nitrite, which cut off the ultra-violet portions of the mercury arc. The excitation was, therefore, only due to the 4358 Å. mercury line, which gave a fairly good and intense spectrum which revealed a large number of even the weaker lines. One anti-Stokes line at frequency 42 cm.<sup>-1</sup> was also observed.

In the molten state the substance got charred within a few hours of exposure, but still a dozen lines have been obtained.

The arrangement of the apparatus was more or less the same as that given by Wood, but the complementary filter method of Ananthakrishnan was not employed since the lines in the vicinity of the exciting mercury line were rather intense. The plate was exposed only for a period of four hours for the low frequency lines, so that the halation due to the  $\lambda$  4358 line was practically nil, giving intense low frequency Raman lines which could be accurately measured.

The glass prism usually employed for photographing such spectra was dispensed with and a Wood's tube was used instead. The substance on

melting lost considerably in volume and hence the tube was filled with molten *p*-diphenyl benzene and allowed to cool slowly. In this way fairly large crystals of the substance were formed in the tube. It was found advantageous to fuse and resolidify the substance in the Wood's tube since a greater quantity of the substance could be used. The transparency of the resolidified mass was also found greater than that in the powdered state. This procedure also considerably reduced the fluorescence, as there were now no interstices left which are bound to be present when the substance is used as a powder. Consequently the incident

TABLE I.  
*p*-Diphenyl Benzene Frequencies.  
(Excitation by Hg.  $\lambda$  4358 only.)

No.	Wave-number	Intensity	in cm. <sup>-1</sup>
1	19896	3	3041
2	20113	0	2824
3	20732	$\frac{1}{2}$	2205
4	20797	$\frac{1}{2}$	2140
5	20842	3	2095
6	20906	3	2029
7	21177	1	1760
8	21264	0	1673
9	21332	10	1605
10	21345	10	1592
11	21388	3	1549
12	21434	1	1503
13	21565	0	1372
14	21663	10	1274
15	21718	6	1219
16	21789	0	1148
17	21898	1	1039
18	21929	4	1008
19	21956	4	981
20	22126	0	811
21	22164	5	773
22	22339	$\frac{1}{2}$	598
23	22852	5	85
24	22895	5	42

light is not multiply reflected at the numerous crystal faces, which possibly adds to the fluorescence already inherent in the substance.

The spectrograms were taken on Agfa isochrome plates which were previously backed, speed H and D. 4400, with a new Füess glass spectrograph having a dispersion of about 21 Å.U. at  $\lambda$  4358. The exposure times varied from 14 to 40 hours. Measurements were made on a new, accurate photomeasuring micrometer. The wave-lengths were calculated, as usual, with the help of Hartmann's linear interpolation formula. Unfortunately no good photograph could be obtained with superposed iron comparison spectrum owing to the presence of intense fluorescence

and the halation due to the stronger lines, which tended also to mask the weaker lines.

The results have been tabulated in Table I. using only the mercury  $\lambda$  4358 radiation. Table II. is a comparative statement of the frequencies

TABLE II.

No.	<i>p</i> -diphenyl benzene (solid state) (Mukerji & Singh)	Diphenyl (solid) (S. A. Aziz)	Diphenyl (solid) (unpublished data) (L. Singh)
1	—	3063 (5)	3065 (4)
2	3041 (3)	3043 (1)	3045 (1)
3	2824 (0)	—	2821 ( $\frac{1}{2}$ )
4	2205 ( $\frac{1}{2}$ )	—	2201 (1)
5	2140 ( $\frac{1}{2}$ )	—	2155 (1)
6	2095 (3)	—	—
7	2029 (3)	—	2035 (1)
8	1760 (1)	—	1774 ( $\frac{1}{2}$ )
9	1673 (0)	—	1680 (0)
10	—	—	1658 (0)
11	1605 (10)	1604 (10)	1610 (10)
12	1592 (10)	1589 (10)	1593 (10)
13	1549 (3)	—	1528 (0)
14	1503 (1)	1506 (2)	1503 (3)
15	—	1459 (1)	—
16	1372 (0)	1326 (0)	1359 (0)
17	1274 (10)	1273 (10)	1278 (10)
18	1219 (6)	1201 (1)	—
19	—	1164 (3)	1166 (2)
20	1148 (0)	1146 (1)	—
21	—	1097 ( $\frac{1}{2}$ )	1089 ( $\frac{1}{2}$ )
22	1039 (1)	1033 (5)	1034 (2)
23	1008 (4)	999 (8)	1003 (5)
24	981 (4)	—	—
25	—	846 ( $\frac{1}{2}$ )	841 ( $\frac{1}{2}$ )
26	811 (0)	—	—
27	773 (5)	784 (2)	783 (1)
28	—	739 (3)	746 (1)
29	598 ( $\frac{1}{2}$ )	606 (2)	615 (2)
30	—	541 (0)	550 (0)
31	—	—	441 ( $\frac{1}{2}$ )
32	—	—	360 (0)
33	—	329 (3)	330 ( $\frac{1}{2}$ )
34	—	—	287 ( $\frac{1}{2}$ )
35	—	243 (2 diff.)	254 (1)
36	—	169 (2 br.)	171 (2)
37	—	146 (4)	150 (2)
38	85 (5)	93 (4 br.)	90 (5)
39	—	54 (3)	55 (5)
40	42 (5)	39 (1)	—

S. A. Aziz, Ind. J. Phys. xiii. part iv. (August, 1939).

diff.=diffuse. br.=broad.

TABLE III.

No.	<i>o</i> -diphenyl benzene (molten state) (Mukerji & Aziz)	<i>m</i> -diphenyl benzene (molten state) (Mukerji & Aziz)	<i>p</i> -diphenyl benzene (solid state) (Mukerji & Singh)
1	3196 (0)	—	—
2	3059 (6)	3062 (2)	—
3	3041 (3)	—	3041 (3)
4	—	—	2824 (0)
5	—	—	2205 ( $\frac{1}{2}$ )
6	—	—	2140 ( $\frac{1}{2}$ )
7	—	—	2095 (3)
8	—	—	2029 (3)
9	—	—	1760 (1)
10	—	—	1673 (0)
11	1608 (5)	1607 (4)	1605 (10)
12	1595 (10)	1597 (10)	1592 (10)
13	1577 (1)	1566 (1)	—
14	—	—	1549 (3)
15	1503 (5)	1494 (4)	1503 (1)
16	1471 ( $\frac{1}{2}$ )	—	—
17	—	1453 ( $\frac{1}{2}$ )	—
18	1430 ( $\frac{1}{2}$ )	—	—
19	—	1403 ( $\frac{1}{2}$ )	—
20	—	—	1372 (0)
21	—	1345 (0)	—
22	—	1309 (10)	—
23	1288 (10)	1279 (2)	1274 (10)
24	1247 (4)	1241 (4 br.)	—
25	—	—	1219 (6)
26	1180 (2)	—	—
27	1158 (6)	1153 (2)	1148 (0)
28	—	1098 ( $\frac{1}{2}$ )	—
29	1059 (2)	—	—
30	1032 (7)	1039 (1)	1039 (1)
31	1005 (8)	1000 (10)	1008 (4)
32	993 (8)	—	981 (4)
33	—	964 ( $\frac{1}{2}$ )	—
34	—	901 (1)	—
35	874 ( $\frac{1}{2}$ )	—	—
36	839 (4)	838 (1)	—
37	—	801 ( $\frac{1}{2}$ )	811 (0)
38	774 (5)	766 (1)	773 (5)
39	744 (2)	—	—
40	708 (5)	707 (3)	—
41	615 (5 br.)	611 (3)	598 ( $\frac{1}{2}$ )
42	558 (3)	—	—
43	521 (2)	—	—
44	501 (1) ?	—	—
45	406 (6)	406 (3)	—
46	359 (5)	—	—
47	319 (3)	—	—
48	253 (1)	275 (3 diff.)	—
49	238 (6 br.)	238 (3 diff.)	—
50	144 (6 br.)	151 (2 diff. & br.)	—
51	112 (5 br.)	—	—
52	73 (5 br.)	80 (4 diff. & br.)	85 (5)
53	—	—	42 (5)

Mukerji and Aziz, 'Nature,' cxlii. p. 477 (September 10, 1938); also Ind. J. Phys. xiii. part 3 (June, 1939).

Mukerji and Aziz, Phil Mag. (7) xxxi. p. 231 (March, 1941).



of diphenyl and *p*-diphenyl benzene. Recent <sup>(4)</sup> results obtained by one of the authors (Laksman Singh) on diphenyl have also been added for the sake of comparison.

Table III. has been prepared with a view to compare the spectra of the ortho, meta and para compounds of the diphenyl benzene family. A comparison of the spectra of these compounds is interesting in so far as the effect of successively loading the carbon atoms in the ortho, meta and para positions by a phenyl group is concerned.

Table IV. gives a comparison of the spectra formed in the solid and molten states. Owing to great experimental difficulties and power-fluorescence only a dozen lines could be obtained in the molten state, and therefore great accuracy in the frequency of these lines cannot be claimed.

One anti-Stokes line corresponding to the shift of frequency 42 recorded above has also been obtained at 43 cm.<sup>-1</sup>.

The spectrogram is reproduced in Plate 000.

TABLE IV.  
*p*-Diphenyl Benzene Frequencies.

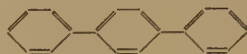
No.	Solid state	Molten state	No.	Solid state	Molten state
1	3041 (3)	3043 (2)	13	1274 (10)	1283 (10)
2	2824 (0)	—	14	1219 (6)	1223 (4)
3	2205 ( $\frac{1}{2}$ )	—	15	1148 (0)	—
4	2140 ( $\frac{1}{2}$ )	—	16	1039 (1)	—
5	2095 (3)	—	17	1008 (4)	1007 (5)
6	2029 (3)	—	18	981 (4)	940 (2)
7	1760 (1)	—	19	811 (0)	—
8	1673 (0)	— (10)	20	773 (5)	761 (2)
9	1605 (10)	1600 (br. & diff.)	21	598 ( $\frac{1}{2}$ )	593 (0)
10	1592 (10)	—	22	—	409 (1)
11	1549 (3)	1523 (2 diff.)	23	85 (5)	93 (3) ?
12	1503 (1)	—	24	42 (5)	—
12 <i>a</i>	1372 (0)	1359 (1)			

In the above Tables intensities are shown in brackets.

### Results and Discussion.

Raman spectrum of diphenyl and of the ortho- and meta-diphenyl benzene have been studied and reported by one of us (Mukerji) and S. Abdul Aziz <sup>(5)</sup> previously. In this investigation we have studied para-diphenyl benzene, a preliminary report on which has already appeared in 'Nature' <sup>(6)</sup>. We can now compare the spectra of these substances and the frequencies of their Raman lines as given in the foregoing tables. We know that in benzene there is a single ring with all the six carbon atoms free. In diphenyl one of the carbon atoms is loaded with a phenyl group. But in ortho- and meta-diphenyl benzene two of the carbon

atoms are loaded with phenyl groups in the ortho and meta positions respectively. In *p*-diphenyl benzene studied here two of the carbon atoms are also loaded with phenyl groups in the para position. The ortho, meta and the para compounds have then three benzene rings and in the para compound these benzene rings are joined end on in the para position.



$p-(C_6H_5)_2 \cdot C_6H_4$ .  
*p*-diphenyl benzene.

The central ring of the para compound then with each of the end ones may be considered to constitute a diphenyl molecule. Diphenyl consists of two benzene nuclei joined end on by a C—C bond in the para direction. It has been shown to be planar. Also all the important frequencies found in diphenyl by one of us <sup>(7)</sup> and also by Wood <sup>(8)</sup> and others <sup>(9)</sup> are observed also in *p*-diphenyl benzene. As Table II. will show, these diphenyl frequencies, also observed in the solid state, are at 3043 (1), 1604 (10), 1589 (10), 1544 ? (Wood), 1506 (2), 1273 (10), 1033 (5), 999 (8), 784 (2), 606 (2) and 93 (4 br.)  $\text{cm}^{-1}$  respectively. It will be seen that in most of the cases the agreement of these frequencies with those of *p*-diphenyl benzene is remarkably close. A comparative study of the spectra observed in these two compounds lends support to the view that the molecule of *p*-diphenyl benzene is planar in form.

In Table III. is given a comparative statement of the frequencies of the ortho-, meta- and para-diphenyl benzene. The most prominent frequencies observed in *p*-diphenyl benzene, as this Table will show, are at 1605, 1592 and 1274  $\text{cm}^{-1}$  respectively. The frequency at 1592  $\text{cm}^{-1}$  is observed in benzene, in diphenyl and also in all the compounds of the diphenyl benzene family. In benzene it is fairly strong; it is stronger still in diphenyl, and is strongest in ortho-, meta- and para-diphenyl benzene. But in symmetrical triphenyl benzene, also studied by the authors <sup>(10)</sup>, it is observed as a strong line at 1603  $\text{cm}^{-1}$ . In fact, there is a close doublet in benzene at frequencies 1584 and 1605. And this doublet occurs very strongly in diphenyl and also in all the three compounds of the diphenyl benzene family. But in triphenyl benzene it is observed only as a single strong line at 1603  $\text{cm}^{-1}$ . This frequency then is characteristic of the benzene ring present in all these compounds and is due to the aromatic linking C=C.

The frequency at 1274  $\text{cm}^{-1}$  observed in *p*-diphenyl benzene is present strongly in all the other compounds of this series, but is not present in benzene. In *m*-diphenyl benzene alone this line is found displaced to 1309  $\text{cm}^{-1}$ , accompanied by a faint companion at 1279  $\text{cm}^{-1}$ . It is evident that the C—C binding amongst the benzene nuclei of these compounds is responsible for the origin of the very strong frequency at 1274  $\text{cm}^{-1}$  observed in *p*-diphenyl benzene.

Para-diphenyl benzene has also shown a fairly strong doublet at frequencies 1008 and 981  $\text{cm}^{-1}$  respectively. In *o*-diphenyl benzene

this doublet is intensely strong, the two frequencies of this doublet being at 1005 and 993  $\text{cm}^{-1}$  respectively. But *m*-diphenyl benzene has shown only one very strong line at 1000  $\text{cm}^{-1}$ , the other component of the doublet not appearing at all on the plate. It may be noted here that the principal Raman line in benzene is at 992  $\text{cm}^{-1}$ , and in diphenyl—one of the three strongest lines—at 1003  $\text{cm}^{-1}$ , the other component of the doublet being entirely absent.

The line at 773  $\text{cm}^{-1}$  observed in *p*-diphenyl benzene also deserves mention. It is observed in all the compounds of this series including benzene and diphenyl. But there is marked variation in the intensity of this line in these compounds. It is weakest in benzene and is slightly stronger in *m*-diphenyl benzene. It is stronger still in diphenyl and is strongest in both ortho- and para-diphenyl benzene. Similar is the case with the line at 598  $\text{cm}^{-1}$  observed in *p*-diphenyl benzene. It is present in all the compounds of this series including benzene and diphenyl. It is present only weakly in *p*-diphenyl benzene but is fairly strong in all the other compounds, with a marked increase in frequency in *m*-diphenyl benzene and a little more in diphenyl and *o*-diphenyl benzene.

*p*-diphenyl benzene has also shown, in the solid state, two extremely low frequency lines, of about equal intensity, at 85 and 42  $\text{cm}^{-1}$  respectively. The line at 85  $\text{cm}^{-1}$  is also observed equally strongly as broad and diffuse bands in ortho- and meta-diphenyl benzene, in the molten state, with a slight decrease in frequency. It has also been observed in solid diphenyl by S. Abdul Aziz<sup>(11)</sup> and by Laksman Singh<sup>(12)</sup> in *o*-diphenyl benzene in the solid state. *p*-Diphenyl benzene, in the molten state, as Table IV. will show, has also shown this line at a slightly higher frequency. This line appears then in practically all these compounds in the solid as well as molten states with slight variations in intensity, and sometimes as a broad and diffuse band in the molten state. Usually lines of such low frequency observed in solids are considered as lattice lines, but as this line is also observed in these substances in the molten state, so, it might be argued, as has been done by Gupta<sup>(13)</sup>, that this low frequency line is due to intermolecular oscillations of polymerized groups and not due to lattice vibrations of the crystal. The line at 42  $\text{cm}^{-1}$  appears in solid *p*-diphenyl benzene as a strong line but it does not appear in the molten compound. Hence this line may be termed as the lattice line.

We will now compare some of the lines observed in molten *p*-diphenyl benzene with those of the solid substance. The frequencies are given in Table IV. It may be mentioned here that in the molten state the substance showed very powerful fluorescence and had a tendency to decompose, and so a good photograph of its spectrum could not be obtained. Great accuracy of the few frequencies obtained cannot also be claimed. But a few interesting comparisons are possible. In the molten state, for instance, the very strong frequency at 1600  $\text{cm}^{-1}$  appears only as a single broad diffuse band of width about 10 Å., instead of a strong doublet which appears in solid *p*-diphenyl benzene and also in the other compounds of the series, including benzene. It is possible that in the molten



state the two lines of the doublet have fused into a single broad band. Symmetrical triphenyl benzene <sup>(14)</sup> has also shown only one strong line at 1603  $\text{cm}^{-1}$  in both carbon disulphide and carbon tetrachloride solutions. This appears then to be the characteristic frequency due to the aromatic linking  $\text{C}=\text{C}$ , and is due to the benzene nucleus present in all these compounds.

The other most prominent frequency observed in molten *p*-diphenyl benzene is at 1283  $\text{cm}^{-1}$ , which is evidently the characteristic frequency due to the  $\text{C}-\text{C}$  bond present in all these compounds.

Molten *p*-diphenyl benzene has given a diffuse line at 1523  $\text{cm}^{-1}$ . It appears that the lines 1549 and 1503 of the solid state have fused into a single line at 1523 in the molten state. The line 1008 of the solid state agrees well with the line 1007 of the molten state in intensity as well as frequency, and there is no change due to change of state. In the solid state there is a fairly strong line at 981  $\text{cm}^{-1}$  and another very weak line at 811  $\text{cm}^{-1}$ . But in the molten state only one, a bit diffuse, line is observed at 940  $\text{cm}^{-1}$ .

The line at 1039  $\text{cm}^{-1}$ , as Table IV. will show, does not appear in molten *p*-diphenyl benzene, although it appears in all the compounds of this series including benzene. Except in diphenyl and *o*-diphenyl benzene it appears as a very weak line in all the other compounds. But it is present in the Raman spectra of all benzene derivatives. It is characteristic of the benzene ring in all these compounds. It is probable, therefore, that this line is completely masked by the very strong fluorescence shown on the plate by this substance in the molten state.

The line at 1359  $\text{cm}^{-1}$  observed in the molten state also deserves attention. In the solid state it is observed at 1372  $\text{cm}^{-1}$  with a slight decrease in intensity. Exactly similar changes are noticed in the line 1318 of molten diphenyl. In the solid state its frequency rises to 1326 and at the same time there is diminution in intensity. The change in intensity appears to be due to the change in state with a slight change in frequency in the same substance. But the big change in frequency from 1326 in diphenyl to 1372 in *p*-diphenyl benzene appears, as already pointed out, to be due to the additional phenyl group in the latter compound in the para direction.

Lastly, it may be pointed out that the lines at frequencies 981 and 811  $\text{cm}^{-1}$  respectively appear to have completely disappeared in the molten state and replaced by a broad line at 940  $\text{cm}^{-1}$ . The principal line in benzene observed at 992  $\text{cm}^{-1}$ , due to a symmetric expansion of the hexagonal ring, has a faint companion at 984 which is definitely attributed to the  $\text{C}_5^{12}\text{C}^{13}\text{H}_6$  molecule and is due to a carbon isotope. The line 981 in *p*-diphenyl benzene seems to correspond with the line 984 in benzene and it may be due to a carbon isotope. In diphenyl this line appears at 980.

A comparative study of the frequencies given in the Tables, II. and III., reveals that practically all the prominent frequencies observed in *p*-diphenyl benzene have also appeared in all the other compounds of this series. These strong common frequencies in the ortho, meta and para



compounds are evidently due to the common benzene nuclei present in these compounds and their existence has already been discussed. But the frequencies which are common only to the ortho and the meta compound are comparatively few. This may be due to their phenyl groups being present in different positions, that is, ortho or meta, in these compounds. But a surprisingly large number of frequencies of solid *p*-diphenyl benzene are found to agree remarkably well with the frequencies of solid diphenyl. A close scrutiny of Table II. will show this.

We have already pointed out that *p*-diphenyl benzene has three benzene rings joined in the para position, and the central ring with each of the end ones may be considered to constitute a diphenyl molecule. It is expected, therefore, that a large number of frequencies present in diphenyl should also be represented in *p*-diphenyl benzene. That this is so is amply verified, as the Table II. will show, by our experiments on *p*-diphenyl benzene. Diphenyl, which consists of two benzene nuclei joined end on by a C—C bond in the para direction has been shown to be planar. Hence these results on *p*-diphenyl benzene lend strong support to the results of L. W. Pickett<sup>(15)</sup> who has made an X-ray study of this compound, that the molecule of *p*-diphenyl benzene is planar in form.

#### References.

- (1) S. K. Mukerji, 'Nature,' cxlii. p. 477 (1938); also Mukerji and Aziz, Ind. J. Phys. xiii. pt. iv. (August, 1938).
- (2) *Loc. cit.*
- (3) Proc. Roy. Soc. A. cxlii. p. 333 (1933); also 'Nature,' cxxxi. p. 513 (1933).
- (4) Unpublished work.
- (5) *Loc. cit.*
- (6) Mukerji and Singh, 'Nature,' cxliv. p. 382 (August 26, 1939).
- (7) *Loc. cit.*
- (8) Wood, Phys. Rev. xxxvi. p. 1431 (1930); xxxviii. p. 2168 (1931).
- (9) Donzelot and Chaix, *Compt. Rend.* cci. p. 501 (1933).
- (10) Mukerji and Singh, 'Nature,' cl. p. 347 (September 19, 1942).
- (11) S. A. Aziz, Ind. J. Phys. xiii. pt. ii. (August, 1939).
- (12) Unpublished work.
- (13) Gupta, Ind. J. Phys. xii. p. 355 (1938); also, x. p. 117 (1936).
- (14) *Loc. cit.*
- (15) *Loc. cit.*

Physics Department,  
Agra College,  
Agra.

#### XCI. Notices respecting New Books.

*Essays in Rheology.* Based on the 1944 Oxford Conference of the British Rheologists' Club. [Pp. vii+163]. (Pitman & Sons, London, 1947.) Price 12s. 6d.

THE word *rheology* was coined by Professor E. C. Bingham in 1928 and defined by him as "a science devoted to the study of the deformation and flow of matter." This definition of the science would seem to be rather wider in scope than is shown by the practice of rheologists, for, as it stands, the definition would include such well-established branches of science as hydrodynamics, aerodynamics, hydraulics and elasticity. In 1928 a Society of Rheology was

founded in the U.S.A., but it was not until 1940 that a more informal, and very lively body, the British Rheologists' Club, was inaugurated in Britain. The Club has held conferences during the years of war which have amply justified its formation, and the volume under review, which illustrates admirably the wide reach of the science, is based on the discussions at a four days' conference held at Oxford in 1944.

As the Scot remarked, "There's fine confused feeding in a haggis," and there is certainly excellent, well-seasoned, and very varied provand in this volume. The titles tell the story: the first essay discusses the rheology of polymers and liquids, and the plasticity of metals; the second, the relation between compression and shear tests; the third, the time-variation of stress and strain; the fourth, nomenclature and symbols; the fifth, rheology in medical science; the sixth, rheology and naval problems; the seventh, and last, rheology in the fine arts. There is much stimulating and important matter in each of these essays, and Dr. Scott Blair and the Club Committee are to be congratulated on the result of their work.

The present reviewer would suggest two topics for the attention of rheologists at some future conference—topics which have received a certain measure of attention, but concerning which the literature is scattered and scanty. The first topic is the capillary flow of liquids under the combined influence of surface tension and viscosity. The problem has been handled by Rideal, Bosanquet and Hartley, Jaeger and Müller, and Washburn. All these deal with newtonian fluids, and the writer is not acquainted with any similar experiments which have been made on non-newtonian fluids.

The second topic concerns the possibility of establishing a coefficient of plasticity and a measure of the yield point of a plastic material by studying the impact of a steel sphere on a slab of plastic or soft metal (*e. g.* lead). It is not difficult to show by experiment that the kinetic energy possessed by the freely falling sphere at the instant of impact is a linear function of the volume of the dent which it makes in the slab of plastic. By tilting the slab slightly so that the sphere bounces and makes a much smaller satellite dent besides the larger or sun dent, we can determine the coefficient of restitution between steel and plastic.

If we take into account the kinetic energy possessed by the rebounding sphere (and neglect any energy which is converted into heat and sound) it is not difficult to determine graphically the volume of the dent produced by the expenditure of unit work, and this would seem to be a reasonable measure for a coefficient of plasticity. Moreover, should it be found that, below a certain height of fall, no appreciable dent is produced, we should have an obvious measure of the yield-point.

There is no end to the problems raised by this fascinating science—a science which provides a modern justification for the saying attributed to Heraclitus, *πάντα ῥεῖ* (all things flow)—a saying which may well serve as the motto of the British Rheologists' Club.

A. F.

*Fourier Transforms and Structure Factors.* By DOROTHY WRINCH. [Pp. 96.] (Asxred Monograph No. 2). (Published by the American Society for X-ray and Electron Diffraction, 1946.) Price \$4.00.

THE help given in analysis of crystal structures by construction of their reciprocal lattices is now well known. The passage from the distribution in the lattice to its reciprocal, and conversely, is a matter of Fourier transforms, and while this is fairly easy for small and simple distributions, it is not so for the more complex ones.

Dr. Wrinch has given here the basic theory, with detailed results for a large number of possible distributions, and with particular attention to cases which are likely to be of value in the analysis of megamolecular crystals. Her book should be a valuable reference work for specialists in this field.

J. H. A.

---

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

## INDEX TO VOL. XXXVII.

- APPLETON (E.) and Hey (J. S.), solar radio noise.—I., 73.
- Auluck (F. C.) and Chowdri (A. G.), a note on fusion and the hole-theory of liquids, 809.
- Baker (G. T.), a bridge method for the investigation of non-linear resistors, 498.
- Band (W.), a critical examination of Milne's kinematical relativity, 551.
- Bending of an initially straight beam under arbitrary loading, 813.
- Beneš (J.) and Petržilka (V.), a method for the determination of crystal cuts by applying the reflection of X-rays from a known lattice plane, 399.
- Bloch (A.), "Lagrangian" formulae for the direct calculation of harmonic output and of intermodulation products, 694.
- Books, new :—The Advancement of Science, Vol. III. No. 12, 72 ; Dr. M. H. Pirene, *The Diffraction of X-rays and Electrons by Free Molecules*, 504 ; R. V. Southwell, *Relaxation Methods in Theoretical Physics*, 576 ; *Advances in Colloid Science*, Vol. II. Scientific Progress in the Field of Rubber and Synthetic Elastomers, 577 ; *Physical Methods of Organic Chemistry*, Vol. II., 577 ; H. Hetényi, *Beams on Elastic Foundation*, 578 ; E. A. Maxwell, *The Methods of Plane Projective Geometry based on the Use of General Homogeneous Coordinates*, 821 ; Prof. C. E. Weatherburn, *A First Course in Mathematical Statistics*, 821 ; *Essays in Rheology*, 883 ; Dr. D. Wrinch, *Fourier Transforms and Structure Factors*, 884.
- Bosworth (R. C. L.), the thermal ohm, farad and henry, 803.
- Boulton (N. S.), plastic stresses in a semi-infinite cohesive mass due to a surcharge covering a strip of the surface of infinite length, 733.
- Boundary perturbations, on the theory of, 563.
- value problems of a heavy circular disk held in a vertical plane, 66.
- Bowden (S. T.) and Jones (W. J.), power laws of latent heat, orthobaric density, surface tension and viscosity of liquids, 480.
- Brown (L. R.), a problem in non-steady heat conduction, 318.
- Campbell (N. R.) and Francis (V. J.), random fluctuations in a cathode ray oscillograph, 289.
- Centrifugal method of measuring the surface tensions and interfacial tensions of liquids, 41.
- Chalmers (J. A.) and Gibby (C. W.) on a difficulty in the application of the law of mass action, 200.
- Chowdri (A. G.) and Auluck (F. C.), a note on fusion and the hole-theory of liquids, 809.
- Colour vision, 33.
- Complete, three-dimensional colour domain and its metrics, 126.
- Conway (H. D.), large deflections of circular and square plates, 756.
- , the large deflections of rectangular membranes and plates, 767.
- Criticism of two well-known theorems on most efficient propellers, 717.
- Crystal cuts, a method for the determination of, by applying the reflection of X-rays from a known lattice plane, 399.
- Crystal habit and internal structure.—I., 184.



Crystal habit and internal structure.—  
II., 217.

—, III., 605.

Davy (N.), the field of a charged semi-infinite rectangular conductor parallel to an earthed infinite plane conductor or the flow from a thick-walled jet, 207.

Deflection of beam, mathematical theory of, 855.

Determination of the orientation of single crystals on a Weissenberg goniometer, 262.

Dielectric properties of liquids in the frequency range 600–3,200 Mc./sec. (50–94 cm.), a method for the measurement of, 747.

Difficulty in the application of the law of mass action, on a, 200.

Diffraction, theory of, 280.

Dingle (H.), a non-quantum indication of Planck's radiation formula, 47.

—, a relativistic determination of the relation between electrical resistance and temperature, 58.

— on the dimensions of physical magnitudes (sixth paper), 64.

Discussion of the pressure-tendencies associated with gradient and horizontal geostrophic flow. A formula for the variation with height of the vertical velocity, 706.

Dunsmuir (R.) and Powles (J. G.), a method for the measurement of the dielectric properties of liquids in the frequency range 600–3,200 Mc./sec. (50–94 cm.), 747.

Dynamic balancing of scanners, a problem in, for radar, 830.

Effect of fluctuation noise interference on pulse distortion, 685.

Electromagnetic field in the neighbourhood of a cusp of a caustic, the structure of, 311.

Electron microscope, the limiting resolving power of the, 677.

Electrosmosis in soils, a laboratory investigation of, 863.

Empirical relations between chemical constitution and thermal expansion coefficient of organic liquids, note on the, 357.

Equations of equilibrium, on a derivation of the, of a thin plate, 648.

Escalator process for the solution of linear simultaneous equations, 106.

Experimental study of the quantum nature of X-rays, 335.

Expression for the sine of a Fourier series, 84.

Fatigue in selenium barrier layer photocells, 13.

Field of a charged semi-infinite rectangular conductor parallel to an earthed infinite plane conductor or the flow from a thick-walled jet, 207.

Finite Hankel transforms, 17.

Forced vibrations of a whirling wire, 426.

Four-figure tables of the Airy function in the complex plane, 236.

Fox (C.), a problem in dynamic balancing of scanners for radar, 830.

Francis (V. J.), theory of the high pressure mercury vapour discharge, 433.

—, —, (B), 653.

— and Campbell (N. R.), random fluctuations in a cathode ray oscillograph, 289.

—, Isaacs (G. G.) and Nelson (E. H.), the luminous efficiency of the high pressure mercury vapour discharge, 789.

Freeman (J. G.), mathematical theory of deflection of beam, 855.

Friend (J. N.) and Hargreaves (W. D.), viscosities of the di-hydroxy benzenes and some of their derivatives, 120.

—, —, the viscosities and structures of acid chlorides and amides, 201.

Fürth (R.) and Pringle (R. W.), a photo-electric Fourier transformer, 1.

Fusion and the hole-theory of liquids, a note on, 809.

Generalized Routh-Hurwitz discriminants, 537.

Gibby (C. W.) and Chalmers (J. A.) on a difficulty in the application of the law of mass action, 200.

Gillies (A. W.) on the bending of an initially straight beam under arbitrary loading, 813.



- Givens (M. P.), an experimental study of the quantum nature of X-rays, 335.
- Ground-water accumulation, Dr. Löwy's theory of, 502.
- Haidinger's rings in divergent light, 700.
- Hargreaves (W. D.) and Friend (J. N.), viscosities of the di-hydroxy benzenes and some of their derivatives, 120.
- , ——, the viscosities and structures of acid chlorides and amides, 201.
- Hey (J. S.) and Appleton (E.), solar radio noise.—I., 73.
- Hilton (P. J.), the effect of fluctuation noise interference on pulse distortion, 685.
- Horner (F.), a problem on the summation of simple harmonic functions of the same amplitude and frequency but of random phase, 145.
- Houstoun (R. A.), fatigue in selenium barrier layer photocells, 13.
- Illusion of size, 643.
- Influence of spray particle size and distribution in the combustion of oil droplets, 94.
- Interference fringes due to a wedge of air or liquid between two glass plates, in theory and practice, 361.
- Interferometry, new contributions to.  
—Part VII. Monochromatic and white-light multiple beam interferometric studies on mica, 390.
- . Part VIII. Interferometric examination of small local irregularities in mica, 453.
- Ionization of mercury vapour by slow electrons, 630.
- Isaaca (G. G.), Francis (V. J.) and Nelson (E. H.), the luminous efficiency of the high pressure mercury vapour discharge, 789.
- Jeans' dynamical theory of gases, 651.
- Jones (W. J.) and Bowden (S. T.), power laws of latent heat, orthobaric density, surface tension and viscosity of liquids, 480.
- Kastner (L. J.), the propulsive force produced by an intermittent jet of gas, 411.
- Katalinić (M.) on the outer unipolar rotation and the outer unipolar induction, 797.
- Keen (B. A.), Dr. Löwy's theory of ground-water accumulation, 502.
- Kinematic relativity, 421.
- , a critical examination of Milne's, 551.
- Kinetic energy correction for the flow of plastic liquids through circular pipes, 571.
- "Lagrangean" formulæ for the direct calculation of harmonic output and of intermodulation products, 694.
- Large deflections of circular and square plates, 756.
- of rectangular membranes and plates, 767.
- Liebmann (G.), the limiting resolving power of the electron microscope, 677.
- Linear network analysis, a contribution to, 778.
- Lower (T. R.), a kinetic energy correction for the flow of plastic liquids through circular pipes, 571.
- Luminous efficiency of the high pressure mercury vapour discharge, 789.
- MacDonald (D. K. C.), a contribution to linear network analysis, 778.
- Maclean (D. J.) and Rolfe (D. W.), a laboratory investigation of electrosmosis in soils, 863.
- Magnetic lens of special field form for  $\beta$ - and  $\gamma$ -ray investigations, designs and applications, 162.
- Martin (D.) on the radial error in a Gaussian elliptical scatter, 636.
- Masket (A. V.), forced vibrations of a whirling wire, 426.
- Mathematical theory of deflection of beam, 855.
- Matthewman (A. G.), a discussion of the pressure tendencies associated with gradient and horizontal geostrophic flow. A formula for the variation with height of the vertical velocity, 706.
- McFadden (T.), the ionization of mercury vapour by slow electrons, 630.
- Measurement of pressure on the seabed, 25.

- Melmore (S.), partitioning of space into enantiomorphous polyhedra, 574.
- Meyerstein (W.) and Morgan (J. D.), a centrifugal method of measuring the surface tensions and interfacial tensions of liquids, 41.
- Millest (D. M.), the sensitivity of a space-charge detector to the presence of positive ions of helium with small kinetic energies, 323.
- Milne's kinematical relativity, a critical examination of, 551.
- Morgan (J. D.) and Meyerstein (W.), a centrifugal method of measuring the surface tensions and interfacial tensions of liquids, 41.
- Morris (J.), an escalator process for the solution of linear simultaneous equations, 106.
- Mukerji (S. K.) and Singh (L.) on the Raman spectrum of *p*-diphenyl benzene, 874.
- Naylor (V. D.), a criticism of two well-known theorems on most efficient propellers, 717.
- Nelson (E. H.), Francis (V. J.) and Isaacs (G. G.), the luminous efficiency of the high pressure mercury vapour discharge, 789.
- Non-linear resistors, a bridge method for the investigation of, 498.
- Non-quantum indication of Planck's radiation formula, 47.
- Oldroyd (J. G.) on a derivation of the equations of equilibrium of a thin plate, 648.
- Outer unipolar rotation and the outer unipolar induction, 797.
- Partitioning of space into enantiomorphous polyhedra, 574.
- Pearcey (T.), the structure of an electromagnetic field in the neighbourhood of a cusp of a caustic, 311.
- Petržilka (V.) and Beneš (J.), a method for the determination of crystal cuts by applying the reflection of X-rays from a known lattice plane, 399.
- Photo-elastic study of stress waves, 463.
- Photo-electric Fourier transformer, 1.  
— mechanism of selenium barrier layer elements, 347.
- Physical magnitudes, on the dimensions of (sixth paper), 64.
- Pidduck (F. B.), theory of diffraction, 280.
- Plastic stresses in a semi-infinite cohesive mass due to a surcharge covering a strip of the surface of infinite length, 733.
- Power laws of latent heat, orthobaric density, surface tension and viscosity of liquids, 480.
- Powles (J. G.) and Dunsmuir (R.), a method for the measurement of the dielectric properties of liquids in the frequency range 600–3,200 Mc./sec. (50–9.4 cm.), 747.
- Pringle (R. W.) and Fürth (R.), a photo-electric Fourier transformer, 1.
- Probert (R. P.), the influence of spray particle size and distribution in the combustion of oil droplets, 94.
- Problem in non-steady heat conduction, 318.  
— on the summation of simple harmonic functions of the same amplitude and frequency but of random phase, 145.
- Propulsive force produced by an intermittent jet of gas, 411.
- Quantum nature of X-rays, experimental study of the, 335.
- Radial error in a Gaussian elliptical scatter, 636.
- Ramachandran (G. N.), Haidinger's rings in divergent light, 700.
- Raman spectrum of *p*-diphenyl benzene, 874.
- Random fluctuations in a cathode ray oscillograph, 289.
- Refractive index of a liquid, determination of, by Newton's rings, 728.
- Relation between some natural constants and the lesser particles, 287.
- Relativistic determination of the relation between electrical resistance and temperature, 58.
- Richardson (E. G.), the measurement of pressure on the sea bed, 25.
- Rolfe (D. W.) and Maclean (D. J.), a laboratory investigation of electrosmosis in soils, 863.

- Sandström (A. E.), the photoelectric mechanism of selenium barrier layer elements, 347.
- Scattering of X-rays of high frequency, 640.
- Searle (G. F. C.), interference fringes due to a wedge of air or liquid between two glass plates, in theory and practice, 361.
- and Woolley (R. L.), determination of refractive index of a liquid by Newton's rings, 728.
- Second law of thermodynamics, with particular reference to Carnot theorems relating to regenerative, refrigeration and warming cycles, 842.
- Sedgwick, (W. F.) on Jeans' dynamical theory of gases, 651.
- Sen (B.), boundary value problems of a heavy circular disk held in a vertical plane, 66.
- Senior (D. A.) and Wells (A. A.), a photo-elastic study of stress waves, 463.
- Sensitivity of a space-charge detector to the presence of positive ions of helium with small kinetic energies, 323.
- Shaxby (J. H.), colour vision, 33.
- Sherman (S.), generalized Routh-Hurwitz discriminant (an extension of the theorems of Sturm, Routh and Hurwitz on the roots of polynomial equations.), 537.
- Siegbahn (K.), a magnetic lens of special field form for  $\alpha$ - and  $\gamma$ -ray investigations, designs and applications, 162.
- Silberstein (L.), the complete, three-dimensional colour domain and its metrics, 126.
- Singh (L.) and Mukerji (S. K.) on the Raman spectrum of *p*-diphenyl benzene, 874.
- Smith (C. J.), further notes and suggestions on the teaching of physics, 505, 579.
- Sneddon (I. N.), finite Hankel transforms, 17.
- Solar radio noise.—I., 73.
- Spectra of the extragalactic nebulae, on the cause of the red shifts in the, 469.
- Steenholt (G.), note on the empirical relations between chemical constitution and thermal expansion coefficient of organic liquids, 357.
- Stokes (A. R.) on a relation between some natural constants and the lesser particles, 287.
- Strachey (C.) and Wallis (P. J.), an expression for the sine of a Fourier series, 84.
- Structure of an electromagnetic field in the neighbourhood of a cusp of a caustic, 311.
- Summation of simple harmonic functions of the same amplitude and frequency but of random phase, problem on the, 145.
- "Tangent-law" photometer, 534.
- Teaching of physics, further notes and suggestions on the, 505, 579.
- Theory of diffraction, 280.
- of the high pressure mercury vapour discharge, 433.
- , — (B), 653.
- Thermal ohm, farad and henry, 803.
- Tolansky (S.), new contributions to interferometry.—Part VII. Monochromatic and white-light multiple beam interferometric studies on mica, 390.
- , Part VIII. Interferometric examination of small local irregularities in mica, 453.
- Viscosities and structures of acid chlorides and amides, 201.
- of the di-hydroxy benzenes and some of their derivatives, 120.
- Walker (W. J.), the second law of thermodynamics, with particular reference to Carnot theorems relating to regenerative, refrigeration and warming cycles, 842.
- Wallis (P. J.) and Strachey (C.), an expression for the sine of a Fourier series, 84.
- Wassermann (G. D.) on the theory of boundary perturbations, 563.
- Weil (R.), an illusion of size, 643.
- Wells (A. F.), crystal habit and internal structure.—I., 184.
- , —, II., 217.
- , —, III., 605.
- Wells (A. A.) and Senior (D. A.), a photo-elastic study of stress waves, 463.



- Whitrow (G. J.) on the cause of the red shifts in the spectra of the extragalactic nebulae, 469.
- Wilson (W.), kinematic relativity, 421.
- Winogradoff (N.), scattering of X-rays of high frequency, 640.
- Woodward (P. M.) and Woodward (A. M.), four-figure tables of the Airy function in the complex plane, 236.
- Woolley (R. L.) and Searle (G. F. C.), determination of refractive index of a liquid by Newton's rings, 728.
- Wooster (N.) and Wooster (W. A.), the determination of the orientation of single crystals on a Weissenberg goniometer, 262.
- X-rays of high frequency, scattering of, 640.
- Yarnold (G. D.), a "tangent-law" photometer, 534.
- Young's modulus, the measurement of, and its temperature coefficient for short filaments by flexural vibrations, 490.
- Yousef (Y. L.), the measurement of Young's modulus and its temperature coefficient for short filaments by flexural vibrations, 490.

END OF THE THIRTY-SEVENTH VOLUME.





[illegible]

DEMCO 38-297